Author:
Jenkins, Jonathan

Title:
Computational design, construction, and characterisation of artificial peroxidases

Computational design, construction, and characterisation of artificial peroxidases

Jonathan Michael Jenkins

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences

School of Biochemistry

September 2018

Word count: 64,869

Abstract

One of the ambitious goals of synthetic biology is to produce biologically compatible catalysts that are cheap, robust, and sustainable. Tailor-made synthetic proteins that can mimic the highly specific and efficient catalysis of natural proteins are key to realising this goal for industrial biotechnology [1]. Oxidoreductases are a diverse class of enzymes that catalyse a similarly diverse array of redox reactions, everything from electron transfer in photosynthesis to site-specific hydrocarbon oxidation in the mammalian liver [2]. Sophisticated oxidoreductase functionality has long been the goal of the maquette protein design approach [3, 4]. The maquette approach aims to understand the necessary sequence structure relationship of these enzymes by iteratively designing in functionality into a non-functional *de novo* scaffold. This bottom up approach to enzyme design aims to produce biocompatible catalysts, desired by the biotechnology industry, whilst simultaneously providing an acid test for our understanding of oxidoreductase enzymology, by building an oxidoreductase from scratch.

This project focuses on *c*-type heme binding maquettes that have been designed to mimic the functionality of peroxidases *en route* to the more commercially interesting oxygenase or peroxygenase chemistry. These *c*-type heme binding maquettes have a *de novo* protein sequence which only contains the essential information for folding into a four-helix bundle that can covalently bind *c*-type heme. This project describes the last design iteration in producing a mono-histidine heme ligation in a *c*-type heme maquette known as C45. C45 functions as a catalytically efficient peroxidase and represents a resounding success of the maquette design principles. Chapter 2 is concerned with the biophysical and kinetic characterisation of the peroxidase activity of C45. Alongside the extensive experimental characterisation this chapter also used computational modelling in the form of: molecular dynamics, small molecule binding screens and *ab initio* protein structure prediction to complement the experimental results. The results of this chapter are an important contribution to our understanding of what the minimal structural requirements to produce native-like peroxidase activity are.

Chapter 3 looks in detail at the reactive intermediates generated during the peroxidase cycle of C45. In particular, the presence of a tryptophan protein radical and its effect on the

functionality of the peroxidase. To do this, a number of tryptophan mutants of C45 were constructed and characterised. The results of this chapter will contribute to future designs of *de novo* proteins that are able to access the reactive heme intermediates of the heme peroxidase and oxygenase reactive cycles.

The fourth chapter compliments the bottom up approach to studying peroxidase activity by using high level computational modelling to study the reactive intermediates of natural peroxidases (CcP and AXP). The dual purpose of these calculations is to both shed light on a contentious issue in the natural peroxidase mechanism and to investigate computational techniques that could be used in the future to inform the design of *de novo* heme protein maquettes. The effect of the distal histidine protonation state on the stability and hydrogen bonding network in the peroxidase active site is the focus of the calculations on the natural peroxidases. These calculations were performed in light of the publication of neutron cryo-crystallography structures that challenged the accepted histidine protonation states of the peroxidase compound I and subsequently the mechanism of proton delivery [5, 6].

This work is evidence of the effectiveness of the bottom up maquette design approach. The minimalist design of the *de novo* peroxidase produces activity that rivals the catalytic efficacy of natural peroxidases. This project also shows the benefit of integrating computational modelling on the *de novo* system to inform the design and characterisation process, as well as the potential to use higher-level calculations of the electronic environment in future *de novo* proteins.

Acknowledgements

I would first like to thank my supervisors Ross and Adrian. I am extremely grateful for all the support and advice they have given me throughout the project and especially during the writing of this thesis. I have thoroughly enjoyed my time in the Anderson Lab and it is largely a result of the supportive and enthusiastic atmosphere that you have cultivated, thank you Ross. I would also like to thank all the other members of the Anderson Lab past and present: Dan, Kris, Bettina, Katie, Angelique, Jack, Rich, George and Claire! In particular, thank you to Dan who took me on as an apprentice and showed me the ropes, and to Katie whose expertise has been greatly appreciated more recently.

I would also like to thank Natalie Fey for her help with the computational chapter as well as: Marc, Richard Lonsdale, Kara, Eric and others in the CCC for their help at various points. I would like to thank Richard Sessions and Amaurys for their generous help and time with setting up molecular dynamics and docking simulations. Many thanks to the High-Performance Computing facility at the University of Bristol for the computational resources used in this thesis. I would also like to thank Muralidharan Shanmugam and the Manchester Institute of Biotechnology for assistance with EPR spectroscopy.

Thanks to all the residence of the Playpen and others on c-floor, for your friendship and sharing many jam scones with me at lunch. Grace Mullally, we start and end on the same day, thanks for your friendship, support and the many pomodoros of writing shared.

I would like to say a special thank you to Mum, Dad, Rach, Ell, Joe, Lucie and the rest of my family and friends who have encouraged me and supported me for much longer than just this project. Thanks to: Dad, Tom, Katie, Nathan, Lois, Aled, Mari and Beth who helped in proof reading various large or small sections of this thesis.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Table of Contents

Abstract **3**
Acknowledgements **5**
Author's Declaration **7**
Table of Contents **9**
List of Abbreviations **12**
List of Figures **15**
List of Tables **17**
1 Introduction **19**
1.1 Protein Folding **19**
1.1.1 Physical code for folding 20
1.1.2 Mechanism of folding 20
1.1.3 Nucleation-condensation mechanism 21
1.1.4 Protein structure prediction 22
1.2 Protein design **24**
1.2.1 *De novo* design 24
1.2.2 Basic principles 25
1.2.3 Coiled-coils 27
1.3 Maquette design approach **28**
1.3.1 Residue helical propensities 28
1.4 Maquette family history **29**
1.4.1 Oxygen binding maquette 31
1.4.2 *c*-type heme maquettes 32
1.4.3 A penta-coordinate heme maquette 33
1.5 Cofactors driving redox chemistry **35**
1.5.1 Electron transfer 36
1.5.2 Heme 38
1.6 Heme oxidoreductases **41**
1.6.1 P450's 42
1.6.2 Intermediates 43
1.7 Heme peroxidases **45**
1.7.1 Substrate interactions 46
1.7.2 Lignin peroxidase 48
1.7.3 DyP-type heme peroxidases 49
1.7.4 Dehaloperoxidase 50
1.8 Classical peroxidase mechanism and kinetics **51**
1.8.1 Formation of Compound I 52
1.8.2 Compound I radicals 55
1.8.3 Compound II 56
1.8.4 Reduction of compound I and II 56
1.8.5 Comparison of Peroxidases and P450s 57
1.9 Computational description of proteins **57**
1.9.1 Molecular Dynamics 58
1.10 Aims **59**
2 Characterisation of the *de novo* peroxidase **61**
2.1 Introduction **61**
2.2 Materials and Methods **63**

2.2.1 Mutagenesis and Cloning 63
2.2.2 Protein expression and Purification 64
2.2.3 UV-Visible Spectroscopy 65
2.2.4 Circular Dichronism 65
2.2.5 NMR 66
2.2.6 Molecular Dynamics 66
2.2.7 Preliminary computational folding with Rosseta 68
2.2.8 Kinetic assays for ping-pong fitting 69
2.2.9 pH-dependent kinetics and spectroscopy 69
2.2.10 Temperature-dependent kinetics 69
2.2.11 Computational binding screen with BUDE 69
2.2.12 Kinetic assays for various substrates 71
2.2.13 Live Cell kinetics 73
2.3 Results 75
2.3.1 Design and mutagenesis 75
2.3.2 Expression and purification 76
2.3.3 UV-visible Spectroscopy 77
2.3.4 CD spectroscopy 78
2.3.5 NMR 79
2.3.6 Molecular Dynamics 81
2.3.7 Preliminary computational folding with Rosetta 84
2.3.8 Kinetic assays for Ping-pong fitting 88
2.3.9 pH Dependence 90
2.3.10 Temperature-Dependent kinetics 92
2.3.11 Computational binding screen 93
2.3.12 Substrate promiscuity 101
2.3.13 Live Cell Assays 115
2.4 Discussion 117
2.4.1 Structure 117
2.4.2 What can be learned from pH dependent activity 118
2.4.3 Proposed mechanism 120
2.4.4 The sticky question of binding 121
3 Studying the *de novo* peroxidase radical 125
3.1 Introduction 125
3.1.1 Radical scavenging 127
3.2 Materials and Methods 128
3.2.1 Mutagenesis and Cloning 128
3.2.2 Protein expression and Purification 128
3.2.3 UV-visible spectroscopy 128
3.2.4 CD 128
3.2.5 Molecular dynamics 128
3.2.6 Mass Spectrometry 129
3.2.7 Electrochemistry 129
3.2.8 Steady-state kinetics assays 130
3.2.9 Total Turnover (TON) 131
3.2.10 Heme degradation assay 131
3.2.11 Intermediate formation 132
3.2.12 Radical scavenging 132
3.2.13 EPR 132

3.3 Results 134
3.3.1 Design and mutagenesis 134
3.3.2 Expression and purification 136
3.3.3 UV Visible spectroscopy 137
3.3.4 CD 139
3.3.5 Molecular dynamics 141
3.3.6 Mass Spectrometry 143
3.3.7 Electrochemistry (OTTLE) 144
3.3.8 Steady-state kinetics with ABTS 145
3.3.9 Turnover Number (TON) 147
3.3.10 Heme degradation assay 147
3.3.11 Intermediate formation 149
3.3.12 Radical scavenging 152
3.3.13 EPR 152
3.4 Discussion 155
4 Computational investigation of peroxidase protonation state 161
4.1 Introduction 161
4.1.1 QM modelling of natural peroxidase systems 162
4.1.2 QM/MM calculations 164
4.1.3 QM cluster models 166
4.1.4 Setup 166
4.1.5 An Engineered Ascorbate Peroxidase 167
4.1.6 Generating MM parameters 169
4.2 Materials and methods 171
4.2.1 Gas-phase quantum mechanics calculations in Jaguar 171
4.2.2 MD and QM/MM 173
4.3 Results 179
4.3.1 QM Cluster Model 179
4.3.2 Iron coordination 179
4.3.3 QM/MM calculations 185
4.4 Discussion 199
4.4.1 QM Cluster Models 199
4.4.2 QM/MM calculations 200
5 Discussion and conclusions 205
5.1 Catalysis 205
5.2 Structure 207
5.3 Radical 208
5.4 QM calculations of natural peroxidases 208
Bibliography 211
Appendix 230

List of Abbreviations

5-ASA	- 5-aminosalicylic acid
ε	- molar extinction coefficient
4ABA	- 4-Aminobenzoic acid
ABTS	- 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)
$ABTS^{\bullet+}$	- 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) radical cation ABTS
AmR	- Amplex Red
AU	- arbitrary units
au	- Hartree atomic units
AXP	- ascorbate peroxidase
AXP2	- an engineered AXP that incorporates N-methyl histidine as the proximal heme ligand
C45	- a maquette based on C46, with the heme C distal heme iron coordinating histidine residue mutated to phenylalanine (H46F)
C46	- a *c*-type maquette with bis-histidine heme ligation
Ccm	- the *E. coli* Cytochrome *c* biogenesis system I
CcP	- Cytochrome *c* peroxidase
CD	- circular dichroism spectroscopy
CI2	- chymotrypsin inhibitor
CL	- cardiolipin
CTM	- *c*-type heme maquette
CV	- column volumes
Cyt *c*	- Cytochrome *c*
cytbc1	- cytochrome bc_1
DAP	- 2,3-diaminophenazine
dARG	- distal arginine conserved in heme peroxidases
DBQ	- 2,6-dichloro-1,4-benzoquinone
DEOB	- 1,4-dimethoxybenzene
dHIS	- distal histidine conserved in heme peroxidases
DHP	- dehaloperoxidase
DMSO	- dimethyl sulfoxide
DNA	- deoxyribonucleic acid
dTRP	- distal tryptophan of some heme peroxidases
DTT	- dithiothreitol
dW2.1 WFF	- W43,80F mutant of C45
dW2.2 FWF	- W8,80F mutant of C45
dW2.3 FFW	- W8,43F mutant of C45
dW3 FFF	- W8,43,80F mutant of C45
dW3Y FFYF	- W8,43,80F F46Y mutant of C45
DyP	- dye-decolorizing peroxidase
ECP	- effective core potential
EDTA	- ethylenediaminetetraacetic acid
Em	- redox transition midpoint potential
EPR	- electron paramagnetic resonance
ESI	- electrospray ionization

FAD/H2 - oxidised/reduced flavin adenine dinucleotide
FMN/H2 - oxidised/reduced flavin mononucleotide
GUA - guaiacol
H2O2 - hydrogen peroxide
HID - histidine protonated on the Nδ only
HIP - doubly protonated histidine
HIS - histidine
HPLC - C18 reversed-phase high-pressure liquid chromatography
HVA - homovanillic acid
IPTG - isopropyl-D-1-thiogalactopyranoside
k_{cat} - maximum rate of an enzymatic reaction
k_{cat}* - maximum rate of an enzymatic reaction determined at rate-limiting peroxide concentrations
K_M - the Michaelis constant, the concentration of substrate at which the rate of an enzymatic reaction occurs at half of k_{cat}
LB - Lysogeny Broth media
LiP - lignin peroxidase
λ_{max} - the absorbance maximum of a UV-visible absorption spectrum
LUM - luminol
MALDI-TOF MS - Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry
MBP - maltose binding protein
MCS - multiple cloning site
MD - molecular dynamics
MM - molecular mechanics
MRE - mean residue ellipticity
MS - mass spectrometry
mV vs NHE - millivolts vs the Nernst hydrogen electrode
NAD/H - oxidised/reduced nicotinamine adenine dinucleotide
NADP/H - oxidised/reduced nicotinamide adenine dinucleotide phosphate
NEB - New England Biolabs
NHE - Nernst hydrogen electrode
NMH - N-methyl histidine
NMR - nuclear magnetic resonance spectroscopy
$OD_{600\ nm}$ - cell growth optical density at 600 nm
oDi - *o*-Dianisidine
oPDA - *o*-phenylenediamine
ORA - Orange II
OTTLE - optically transparent thin-layer electrochemistry
P450 - cytochrome P450
PAA - peracetic acid
PABA - *p*-aminobenzoic acid
pAN - *p*-anisidine
pASP - proximal aspartic acid of heme peroxidases
PDA - photodiode array detector
PDB - Protein Data Bank

pEC86	- pEC expression vector with the *E. coli* Cytochrome *c* biogenesis system I proteins A-H cloned into the multiple cloning site
pHIS	- proximal histidine of heme peroxidases
pK_a	- the pH at which a specific protonatable molecule is 50% protonated and 50% deprotonated in solution
pTRP	- proximal tryptophan of heme peroxidases
Q-bands	- 450-600 nm visible region of heme protein absorption
QM	- Quantum Mechanics
QM/MM	- hybrid quantum mechanics/molecular mechanics calculations
RMSD	- root mean squared deviation
ROS	- reactive oxygen species
RPM	- revolutions per minute
SDS-PAGE	- Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SHE	- standard hydrogen electrode
TBP	- 2,4,6-tribromophenol
TCA	- tricarboxylic acid cycle
TCEP	- tris(2-carboxyethyl)phosphine
TCP	- 2,4,6-trichlorophenol
TEMPO	- 2,2,6,6-tetramethylpiperidinyl-1-oxy
TEV	- Tobacco Etch Virus
TFA	- 2,4,6-trifluoroacetic acid
Tm	- thermal melting transition midpoint temperature
TON	- turnover number, moles of substrate turned over mole of enzyme
UOB	- University of Bristol
UV	- ultraviolet
V0	- initial reaction velocity (concentration increase per unit time) of an enzymatic reaction
VA	- veratrole alcohol (veratryl alcohol)
VP	- versatile peroxidase
WT	- wild type

List of Figures

Figure 1.1.1 Protein folding mechanisms 22
Figure 1.2.1 Comparison of new folds and protein structures over time 25
Figure 1.2.2 Helical Wheel diagram 27
Figure 1.4.1 Maquette Family History 30
Figure 1.4.2 Heme B and C comparison. 33
Figure 1.4.3 Maquette design cycle. 34
Figure 1.5.1 Redox cofactors 36
Figure 1.5.2 Redox midpoint potentials compared. 38
Figure 1.5.3Heme modifications 40
Figure 1.6.1 Reaction cycle of P450's and peroxidases 44
Figure 1.8.1 Compound I formation mechanisms 53
Figure 1.10.1 MM energy terms. 59
Figure 2.3.1 Schematic of C46 and C45 75
Figure 2.3.2 SEC and SDS-PAGE results for C45 76
Figure 2.3.3 Maquette UV- visible spectra. 77
Figure 2.3.4 CD results for C45 78
Figure 2.3.5 CD spectrum od C45 with urea 79
Figure 2.3.6 NMR of C45 80
Figure 2.3.7 C45 MD comparison. 82
Figure 2.3.8 MD structures of C45. 83
Figure 2.3.9 Rosetta structures with heme. 85
Figure 2.3.10 Rosetta structures without heme. 87
Figure 2.3.11 ABTS reaction with C45 88
Figure 2.3.12 Ping-pong kinetics C45 89
Figure 2.3.13 pH dependent steady-state kinetics 91
Figure 2.3.14 Temperature-Dependent steady-state kinetic 92
Figure 2.3.15 BUDE process illustration 93
Figure 2.3.16 Binding Screen Clusters 95
Figure 2.3.17 Structures of small molecule ligands 98
Figure 2.3.18 Binding screen result. 99
Figure 2.3.19 Binding screen results continued. 100
Figure 2.3.20 Cytochrome c oxidation with C45. 102
Figure 2.3.21 o-Phenylenediamine oxidation with C45. 103
Figure 2.3.22 Amplex Red oxidation with C45 104
Figure 2.3.23 Guaiacol oxidation with C45. 105
Figure 2.3.24 Luminol oxidation with C45 106
Figure 2.3.25 o-Dianisidine oxidation with C45 107
Figure 2.3.26 TCP oxidation with C45 108
Figure 2.3.27 TBP oxidation with C45 109
Figure 2.3.28 Substrates without steady-state kinetics. 110
Figure 2.3.29 Substrates with very poor or no turnover. 111
Figure 2.3.30 C45 of kinetics results. 113
Figure 2.3.31 Live cell Assays. 116
Figure 2.4.1 C45 Mechanism. 121

Figure 3.1.1 Tryptophan positions in C45 model. 127
Figure 3.3.1. Design of tryptophan knock-out maquettes 135
Figure 3.3.2 SEC and SDS-PAGE results for dW mutants. 136
Figure 3.3.3 UV-Visible spectroscopy of dW mutants. 137
Figure 3.3.4 Ferric 280 nm signal and ferrous 430 nm shoulder heights 138
Figure 3.3.5 CD results comparison for dW mutants. 139
Figure 3.3.6 CD thermal melt spectrum scans for dW mutants and C45 140
Figure 3.3.7 MD structures. 141
Figure 3.3.8 Averaged MD data.. 142
Figure 3.3.9 Mass spectra. 143
Figure 3.3.10 OTTLE results.. 145
Figure 3.3.11 Steady-state results for dW mutants.. 146
Figure 3.3.12 Turnover number dW mutants with ABTS. 147
Figure 3.3.13 Soret peak changes on mixing with H_2O_2. 148
Figure 3.3.14 Intermediate UV-visible spectra 150
Figure 3.3.15 Intermediate spectra comparison 151
Figure 3.3.16 Rate of formation of reaction intermediate 151
Figure 3.3.17 Wide sweep EPR spectra 153
Figure 3.3.18 EPR spectra annotated 153
Figure 3.3.19 A comparison of radical signal 154
Figure 4.1.1 Naming convention for histidine tautomers in AMBER. 161
Figure 4.1.2 Compound I formation mechanisms.. 162
Figure 4.1.3 QM/MM calculation setup. 165
Figure 4.1.4 Neutron structure active site conformation. 167
Figure 4.1.5 AXP2 active site conformation 168
Figure 4.1.6 CHARMM atom naming convention 170
Figure 4.3.1 Structural comparison of results with QM1-Di setup 180
Figure 4.3.2 Distal hydration from QM1-Di minimised structures 183
Figure 4.3.3 MD Hydration plots 187
Figure 4.3.4 Hydration and RMSD plots 188
Figure 4.3.5 Iron oxygen and iron nitrogen bond distances 190
Figure 4.3.6 QM/MM energy differences to move from dHIP to dHID 191
Figure 4.3.7 QM/MM result distance measures 192
Figure 4.3.8 CcP QM/MM optimised structures 193
Figure 4.3.9 AXP QM/MM optimised structures.. 194
Figure 4.3.10 AXP2 QM/MM optimised structures.. 195
Figure 4.3.11 Spin density on heme and pTRP 197
Figure 4.3.12 Atomic charges and spin density on heme Fe and oxyferryl-O 198
Figure 4.4.1 Suggested peroxidase mechanism.. 203

List of Tables

Table 2.3.1 pH dependence Michaelis–Menten fit values. 91
Table 2.3.2 C45 pH dependent midpoint potentials (E_M). 92
Table 2.3.3 Temperature dependence Michaelis–Menten fit values. 93
Table 2.3.4 Comparison of kinetics results. 113
Table 2.3.5 Oxidation Potentials 114
Table 3.1.1 Tryptophan knock-out mutants. 126
Table 3.3.1 Tryptophan knock-out genes produced. 134
Table 3.3.2 Deconvoluted UV visible peaks. 138
Table 3.3.3 MADLI mass peak compared to predicted mass. 144
Table 3.3.4 Comparison of dW mutant kinetic results. 146
Table 3.3.5 Heme degradation rates 148
Table 3.3.6 Q-band intermediate peak positions for dW mutants. 150
Table 4.2.1 QM cluster calculations structures. 171
Table 4.2.2 Structures used for MD and QM/MM geometry optimisation. 173
Table 4.3.1 Difference in energy between QM cluster structures. 181
Table 4.3.2 Distance information from QM1-Di calculations 184
Table 4.3.3 Charges on central heme atoms. 186

Chapter 1

Introduction

The defining test of our understanding of any natural system is our ability to recreate it. As famously noted by Richard Feynman, "What I cannot create, I do not understand" [7]. To this end, the field of *de novo* enzyme design seeks to understand the relationship between protein sequence, structure and function in enzymatic catalysis by recreating it. Proteins are the functional molecules of life and perform an astonishing array of chemical processes, highly specifically and generally in ambient conditions. Proteins as nanoscale machines facilitate everything from the capture of solar energy and storage in various forms, to carefully curating, reading and replicating DNA. The potential uses for protein-based materials, systems and therapies to address many fundamental challenges facing mankind is immense.

This project aims to better understand the necessary components needed to produce sophisticated oxidoreductase activity in a *de novo* protein. This will be primarily achieved using the bottom-up maquette design approach. Computational modelling on natural peroxidases will also be performed to further our understanding of the peroxidase reactive mechanism and to set the context for using same modelling techniques in the *de novo* system in the future.

1.1 Protein Folding

Our understanding of protein folding is founded on the success of structural biology in producing atomic resolution structures of proteins in native-like conformations. The first structure of a globular protein in 1958 set in motion the huge biophysical conundrum that is the protein folding problem [8]. The level of complexity and lack of symmetry which astounded Kendrew *et al*. with their initial structure of myoglobin has been observed again and again with successive protein structures solved. Currently there are over 60 thousand unique atomistic protein structures solved but work on a solution of the protein folding problem is still very much in progress [9]. The protein folding problem has often been described by two key questions [10]. First, what is the physical code, what physical and chemical processes allow the three-dimensional structure of a protein to be fully expressed from its one-dimensional sequence? Second, what is the folding mechanism that reduces the

terrifyingly vast number of possible conformations even a short protein can adopt, and how does this process occur so quickly? The famous calculation by Cyrus Levinthal showed that this clearly could not be a random process of sampling all possible conformations [11]. To illustrate the point, we can construct a model protein and vastly restrict its conformational flexibility to just 10 possible values of the φ and ψ angles in the backbone, and allow each conformation to be sampled as fast as a single bond rotation (10^{-13} s) [12]. Under these conditions a protein of just 32 amino acids would take 31 billion years to sample all possible conformations, which is universally agreed to be too long.

1.1.1 Physical code for folding

The driving force for protein folding is a combination of physicochemical factors. Hydrogen bonds play a key role, which was first shown in the stabilization of secondary structure and predicted to be a key interaction in helix promotion as far back as 1951 [13]. Van der Waals and electrostatic interactions clearly play an important role in such tightly packed molecules. The protein backbone angle preferences add an additional restriction to folding. However, hydrophobic/hydrophilic interactions between the protein and solvent seem to be the largest driving force of protein folding [14].

1.1.2 Mechanism of folding

In terms of a mechanism, protein folding is now understood as a landscape, which for natural proteins is shaped like a funnel. This funnel consists of many higher-energy unfolded conformations at the upper edges. As the protein folds into lower-energy conformations there are fewer and fewer possible globular structures at the center of the funnel [10]. In this funnel, adding an energy bias to correct folding on the order of just 2kT (where k is Boltzmann's constant and T the temperature) can reduce the time of a random folding search of a protein to the order of seconds [15]. At the bottom of this funnel is the native state of a protein, which is described as the global minimum Gibbs free energy conformation formalized in 1973 with Anfinsen's "thermodynamic hypothesis" [16]. The description of a protein folding landscape does not give a mechanistic description of how the protein will fold, only why it does from a thermodynamic perspective. Many different folding mechanisms have been postulated. The **Nucleation-growth mechanism** was one of the first to be rejected. It hypothesizes that tertiary structure propagates out rapidly from a nucleus of secondary

structure without any folding intermediates. The isolation and study of protein folding intermediates in the 1970's made this hypothesis unpopular [17].

Two other proposed mechanisms were the framework mechanism and hydrophobic collapse (see Figure 1.1.1) [18, 19]. The **framework** mechanism sees the secondary structure fold, first with the tertiary structure, then buildup by the docking of completed secondary elements. This mechanism was born out of studies on small stable helical peptides that refolded rapidly after denaturation [19]. These secondary structural elements, that were stable in solution, were then thought to form a framework that the tertiary structure was built around. However, it is relatively rare to find such strong structural preferences in normal protein sequences. Additionally, the majority of peptides that make up secondary structure are disordered when taken out of context [20]. Long range interactions between residues sequentially far apart, were also seen to be important in producing and stabilizing secondary structure [21]. All three observations are in contradiction to a purely framework mechanism.

In the **hydrophobic collapse** model, folding is initiated by coalescing of the hydrophobic residues and the expulsion of water from the core [18]. In this smaller volume, proteins would then have a confined conformational search for the native structure. Early support came from studies showing that hydrophobic interactions were the major contributor to protein stability for most proteins [22]. But, questions were raised over if it would be possible to overcome the reorganizational energy needed from secondary structure within a nonspecific hydrophobic globule.

1.1.3 Nucleation-condensation mechanism

It is now thought likely that most proteins do not fold by either extreme (framework or hydrophobic collapse) but through some sort of hybrid. This middle ground is sometimes termed the Nucleation-condensation mechanism. Fink *et al*. analyzed structural properties of 41 native and partially-folded (occupying a stable intermediate) proteins trying to address what they term the "chicken-egg scenario" of condensation and secondary structure formation [23]. They observed a good correlation between secondary structure formation alongside deceasing hydrodynamic volume. The discovery of proteins like chymotrypsin inhibitor (CI2) that fold with two-state kinetics without accumulating folding intermediates allude to the possibility of the Nucleation-growth mechanism [24]. Φ-value analysis of CI2 indicated the secondary and tertiary structure formation occur in parallel with hydrophobic collapse [25].

Φ-value analysis uses point mutations to probe different amino acids' contribution to the native folding kinetics. These techniques have allowed the following general principles from small globular protein folding to be proposed, and could be abstracted to most proteins. First, folding is initiated by the formation of a nucleus. This nucleus is made up of consolidation of the hydrophobic core concurrent with secondary and tertiary structure formation. This folding nucleus often contains a high proportion of secondary structure. Second, the formation of the nucleus involves a condensation of the molecule that is largely driven by hydrophobic effects. Third, early folding stages can involve multiple pathways but the latter stages are more unified in the conformational path taken to the native structure [26].

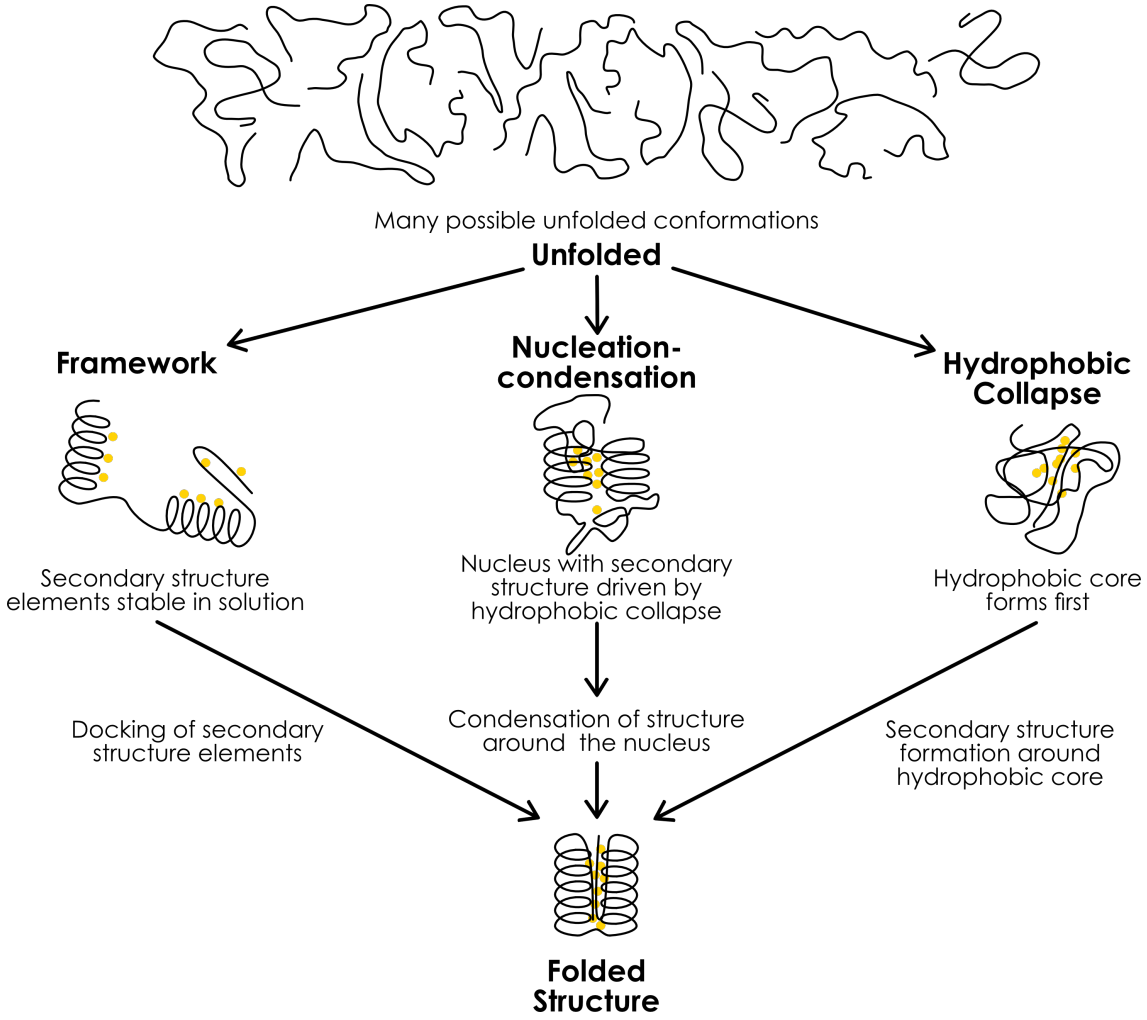


Figure 1.1.1 Protein folding mechanisms. *The yellow spheres represent hydrophobic residues and coils represent secondary structure.*

1.1.4 Protein structure prediction

With developments in genomics, the thousand-fold gap between the number of protein sequences we know and the number of structures available is growing [27]. This problem is

partially addressed by methods of homology modeling which can predict the structure of homologous proteins that have high sequence identity. Where this is not the case, other computational methods are needed. Critical Assessment of Structure Prediction (CASP) meetings were setup to experimentally test these methods of protein structure prediction and their results demonstrate gradual improvements every two years since its start in 1994 [28]. CASP tests structure prediction algorithms on sequences that have not yet been made publicly available, allowing the predictions to be scored against the experimental structure. All the successful algorithms rely heavily on the PDB as they assume that similar sequences will have similar structures. HHpred is an interactive server for protein homology detection and structure prediction that consistently performs well in CASP. It is purely a fragment-based method [29].

Rosetta is another protein structure prediction software that is leading the field in *ab initio* structure prediction or free modeling, where no high identity homology structure for the target exists [30]. Rosetta is a hybrid approach that combines small fragments of structural homology derived from the PDB (either three or nine amino acids long) and all-atom structural refinements. In general, Rosetta *ab initio* protocols first fit the short fragments on to the sequence. Then using a Monte Carlo search protocol, perturbations are made to the proteins with the side chains represented as centroids. A scoring function composed of terms representing close range physical interactions (hydrogen bonding, hydrophobic burial and electrostatics) and sequence dependent terms (assembly of α and β structure) selects for folded results. A second all-atom refinement step then takes place using the Rosetta full-atom force field. In CASP10 around 10% of targets were predicted accurately enough to guide biochemical, industrial study or be used as an experimental native structure could [10, 28]. An additional challenge is that intrinsically disordered proteins may account for a large fraction of the eukaryotic proteome, with over 30% of human proteins predicted to be fully or partially disordered [31]. To predict the dynamics and structures of these proteins, more work on understanding and modeling the fundamental physical force driving folding is required. This will build a more general, physical description of folding and reduce our reliance on existing protein structural databases that are missing these types of proteins.

1.2 Protein design

For many groups in more recent history, protein structure prediction algorithms and protein design have gone hand in hand. This is most clearly seen in the use of Rosetta to remodel and produce new protein structures and even enzymes. The most notable result in this case was the design of a protein fold that had not been found in nature, Top7 [32]. The design process used a *de novo* backbone sketch as a starting structure to fit PDB fragments to, creating a large backbone ensemble. This ensemble then went through rounds of sequence optimization and backbone refinement scored with the Rosetta force field.

A step change in the utility of this software was its use in designing Kemp elimination functionality in to an existing protein scaffold [33]. Here Rosetta's packing and scoring algorithms were utilized to redesign an existing natural protein structure to bind and stabilize a theoretical transition state of a chemical reaction. The initial result from the computational library (WT) had modest activity that was improved 200-fold after 7 rounds of directed evolution (R7). The catalytic efficiency (k_{cat}/K_M) of the WT was 12 $M^{-1}s^{-1}$and after directed evolution R7 had a k_{cat}/K_M of 2590 $M^{-1}s^{-1}$. This Rosetta design process was also able to produce enzymes that catalyse retro-aldol and Diels–Alder reactions with both initial structures also greatly improved though directed evolution [34, 35]. Whilst this constitutes a watershed moment for computational design it is not the start of *de novo* protein design. In fact some of the early protein designers (DeGrado *et al.*) were able to produce a Kemp eliminase with activity comparable to that of the initial Rosetta design before directed evolution, just using their chemical intuition [36].

1.2.1 *De novo* design

Current computational methods rely heavily on the PDB with the results of these methods biased towards looking like other experimentally determined structures. A greater challenge is to use a fully *de novo* protein scaffold as a starting point for designing activity. But why would anyone punish themselves in this way? There are several good reasons. Natural enzymes are the result of a complex mix of evolutionary pressures through time. Many of these selective pressures have little to do with the primary enzymatic activity and more to do with the protein existing in a dense soup of cellular interactions. From an explorative viewpoint, studying the fundamental principles that drive activity in a system this free from evolutionary complexity gives a unique and invaluable perspective.

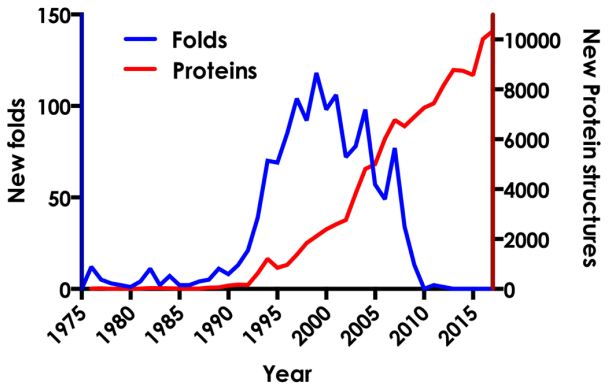


Figure 1.2.1 Comparison of new folds and protein structures over time. *The blue line and left axis show the number of new basic protein folds added to the CATH database per year. The red and right y axis show the number of protein structures deposited on the PDB per year.*

Another good reason is that nature only uses a very small area of protein sequence space, and the trend of discovery of new protein folds suggests that the same seems to be true for protein structure space. There have been no new basic folds listed in the CATH database (cathdb.info CATH v4.0) since 2012, while the number of new proteins structures deposited on the PDB has been steadily increasing per year from 7261 in 2010 to 10334 in 2017 [9]. *De novo* protein design could then be the only way to discover new and perhaps better protein structures.

A third reason for *de novo* design is to more easily create proteins and enzymes orthogonal to natural pathways when used or expressed *in vivo*. These components designed to have few or no interactions with natural proteins could play an important role in constructing and understanding synthetic genetic circuits and other synthetic pathways. One final advantage of *de novo* proteins and enzymes is to construct proteins robust enough to cope with the difficult conditions needed for industrial use. Large free energies of folding are unlikely to have been a strong driving force in natural protein evolution as a free energy of folding of 8 kcal.mol^{-1} corresponds to having over 99.999% of the population folded [37]. Through designing and understanding optimal protein folding it will be possible to make proteins more stable than is required by nature.

1.2.2 Basic principles

Some of the very first *de novo* proteins were designed in the late 80's. Both DeGrado *et al*. and Hecht *et al*. separately produced *de novo* four-helical bundle proteins by using the general principles of protein folding understood at the time. Hecht *et al*. aimed at building a

native like sequence using almost the full complement of amino acids [38]. DeGrado *et al*. aimed to understand the minimal sequence requirements for producing this tertiary structure so used as few amino acids as possible [39, 40]. This early work coevolved with an increasing number of studies aiming to understand the thermodynamic contributions of different amino acids to produce α-helical and β-strand structures [41, 42]. The design process took an iterative approach, testing the effect of each change on the structure. The designed proteins were built with hydrophobic interactions as the driving force for folding. This was achieved by using binary patterning of hydrophobic (**H**: Ala, Phe, Ile, Leu, Met, Val, Trp, and Tyr) and polar (**P**: Asp, Glu, His, Lys, Asn, Gln, Arg, Ser, and Thr) residues to build up a core and surface, burying H and exposing P sidechains [43]. The most naive patterning would comprise an alternating repeat (HPHP…) along the polypeptide chain. This type of repeat lends itself to forming **β-structure**. In a β-sheet alternate sidechains point in opposite directions allowing two or more stands to pack together, burying the hydrophobic sidechains [12].

α-helical structure corresponds to right-handed helices where backbone carboxyl groups (residue n) hydrogen bonds along the helical axis of the amino group of the residue four along (n+4). Each turn of an α-helix has a pitch (the rise along the helical axis) of 5.4 Å and is made up of 3.6 residues. In most proteins α-helices are on average 12 residues or 18 Å long [13]. The residue sidechains point outward from the helix and can be represented on helical wheel diagrams (Figure 1.2.2). These diagrams can be used to aid protein design. The position hydrophobic residues along one face of the helix can produce a core when one or more amphipathic helices combine, as seen in Figure 1.2.2.B and C.

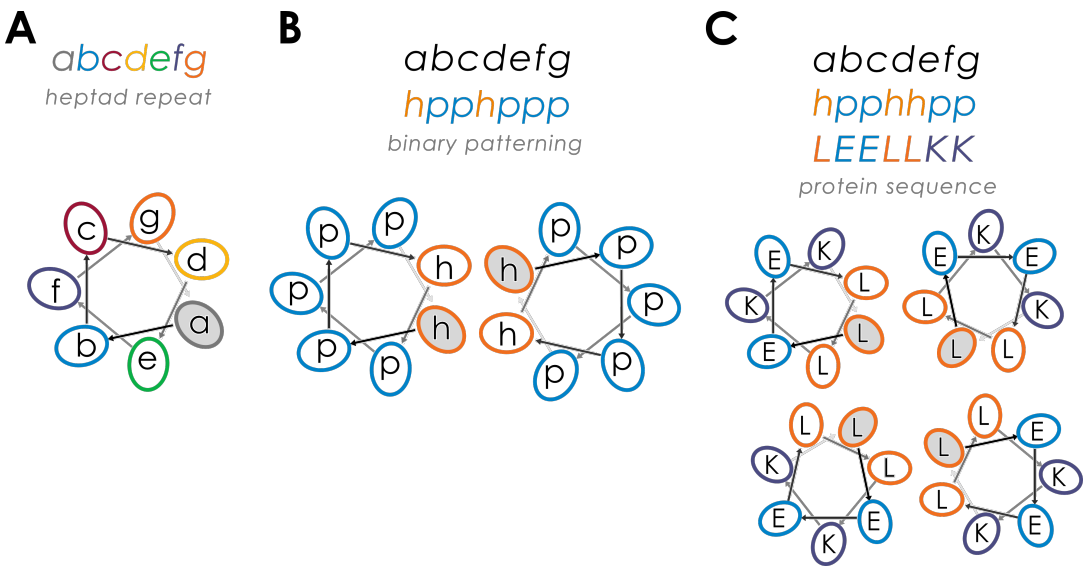


Figure 1.2.2 Helical Wheel diagram.
(A) Shows a right hander α-helical heptad repeat. The letters a to g are given to the positions of the heptad. The direction of the oval at each position show the direction the side chain protrudes out of the helix. (B) Shows the binary patterning of a coiled-coil. "h" are hydrophobic residues and "p" are polar residues. (C) Shows a helical wheel representation of the protein sequence of the first maquette four-helical bundle (see Figure 1.4.1.1).

1.2.3 Coiled-coils

This hydrophobic patterning and core packing have been well studied in natural and *de novo* coiled-coils. Canonical coiled-coils have a tandem heptad repeat that form amphipathic α-helices which assemble into helical bundles with left handed supercoils [44]. The heptad repeat in its binary form is generally HPPHPPP and is given the notation *abcdefg* for the 7 positions around the helix where *a* and *d* are hydrophobic (Figure 1.2.2.B). Since 7 residues are not quite 2 turns ($3.6 \times 2 = 7.2$) an imperfect hydrophobic interface is created, resulting in supercoiling. The core packing in coiled-coils is stricter than in non-coiled-coil 4 α-helical bundles; they obey knobs-in-holes packing first predicted by Crick in 1953 [45]. Differing *a* and *d* hydrophobic residues will have preferred packing arrangements. Changing between leucine isoleucine and valine (L, I and V) in these positions produces different oligomerisation states. The *a* and *d* position of dimers are generally V/I and L or , trimers I and I and tetramers L and I [46]. Positions *e* and *g*, that often make up the interface are between oligomers, are important in producing polar interactions that enhance the specificity for particular oligomerisation states. This parameterisation of coiled-coil structure has allowed protein designers to produce an array of different complexes and barrels, many of which are not observed in nature [47].

1.3 Maquette design approach

A maquette is a term borrowed from sculpting and architecture. It is a small, cut-down model built before a full sculpture is created with the essential features of the design but in a simple form. Protein maquettes are small *de novo* protein scaffolds that are intended to aid our understanding of the minimal requirements of specific sequence-structure-function relationships. Generally, maquettes have been designed to have 4 α-helical bundle tertiary structure. 4 α-helical bundles are simple enough to be produced and studied as *de novo* proteins but robust enough to incorporate asymmetry and cofactor binding. These maquettes have their evolutionally roots back in the 1980's with Degrado *et al.* [40]. Each iteration of design changes made to these proteins is tested by producing and biophysically characterising any variation from the previous generation.

A theme of maquettes design is a drive for functionality over form (atomistic structural information). This thirst for useful activity has led to the incorporation of cofactors into maquettes. Particularly for redox chemistry, the incorporation and redox active cofactors is the best way to impart functionality. Then the exploration to find the minimal requirements need to produce a desired activity can be performed around the cofactor, the centre of and otherwise beige canvas [48]. This is a bottom-up approach to studying enzymology which complements classical studies on natural systems.

1.3.1 Residue helical propensities

Four helix bundles (4HB) are a relatively common and well-studied tertiary structure. Parameterisation of these structures has been led by insights from published experimental structures as well as studies on folding, stability of helix formation, topologies available and cofactor binding [49–51]. Deciding which residues to pattern the helices of the 4HB is made with attention to the individual amino acid's particular propensity for forming alpha helices. Glycine, proline, serine, aspartic acid and asparagine have low α-helical propensity. Glycine is small and too flexible while the heterocyclic side chain of proline excludes the backbone angles needed for helix formation. Proline is therefore considered a helix breaking residue. Serine, aspartic acid and asparagine have short and polar side chains that can disrupt the backbone hydrogen bonding. Valine and isoleucine are also generally disfavoured for helix formation as they branch at the C_β the first carbon of the side chain which hinders α-helix formation. The maquette sequence is therefore primarily comprised of amino acids that favour

helix formation: glutamate (E), alanine (A), leucine (L), glutamine (Q), lysine (K) and arginine (R) [52].

The net charge of most maquettes is around -20 (Figure 1.4.1-C45 is -18 in the ferrous form). This supercharged exterior serves to prevent oligomerisation and protect against aggregation. Studies comparing positive and negative supercharged Green fluorescent protein (GFP) with wtGFP, demonstrated that supercharging the protein decreased aggregation and a greater proportion of supercharged GFP retained its fluorescence characteristics after denaturation [53].

1.4 Maquette family history

Maquettes have been designed to incorporate many natural cofactors including: heme, chlorins, flavins and quinones [54–58]. For more information on cofactors see Section 1.5. They have been used to study allosteric and charge activated conformational switching, cofactor binding, electron transfer and oxygen binding [59–62]. Heme A and B are the most common cofactors studied within 4HB maquettes. This section aims to chronicle the ancestry of a heme-peroxidase maquette which is the focus of this project.

Hydrophobic collapse is considered to be the driving force for folding the 4HB in solution. Soluble maquettes use a heptad HPPHHPP binary pattern to form amphipathic helices that combine, forming a hydrophobic core. Unlike *de novo* coiled-coils the core packing interaction of 4HB are not specifically defined. The **first** generation maquettes were symmetrical helices that were synthesised by solid state peptide synthesis and oligomerised into a 4HB [39]. They were short, just two near-identical heptad repeats, and were composed of only three different amino acids (LEELLKKLEELLKL). To incorporate a cofactor, the short minimal sequence was lengthened to the size of a typical, natural 4HB. The result is a bundle where each helix is slightly more than three heptads long (27 residues) and can exist in a parallel or anti-parallel topology. The **second** generation added a N-terminal CGGG sequence which could restrict the topology to *syn* or *anti* through disulphide-mediated dimerization of the cysteine (C) side chains. One leucine on each helix was also replaced with a tryptophan (W) to improve optical detection of the peptides. The loop regions of maquettes are generally composed of glycine (G) and serine (S) as these residues have particularly low helical propensity and a high degree of conformational flexibility.

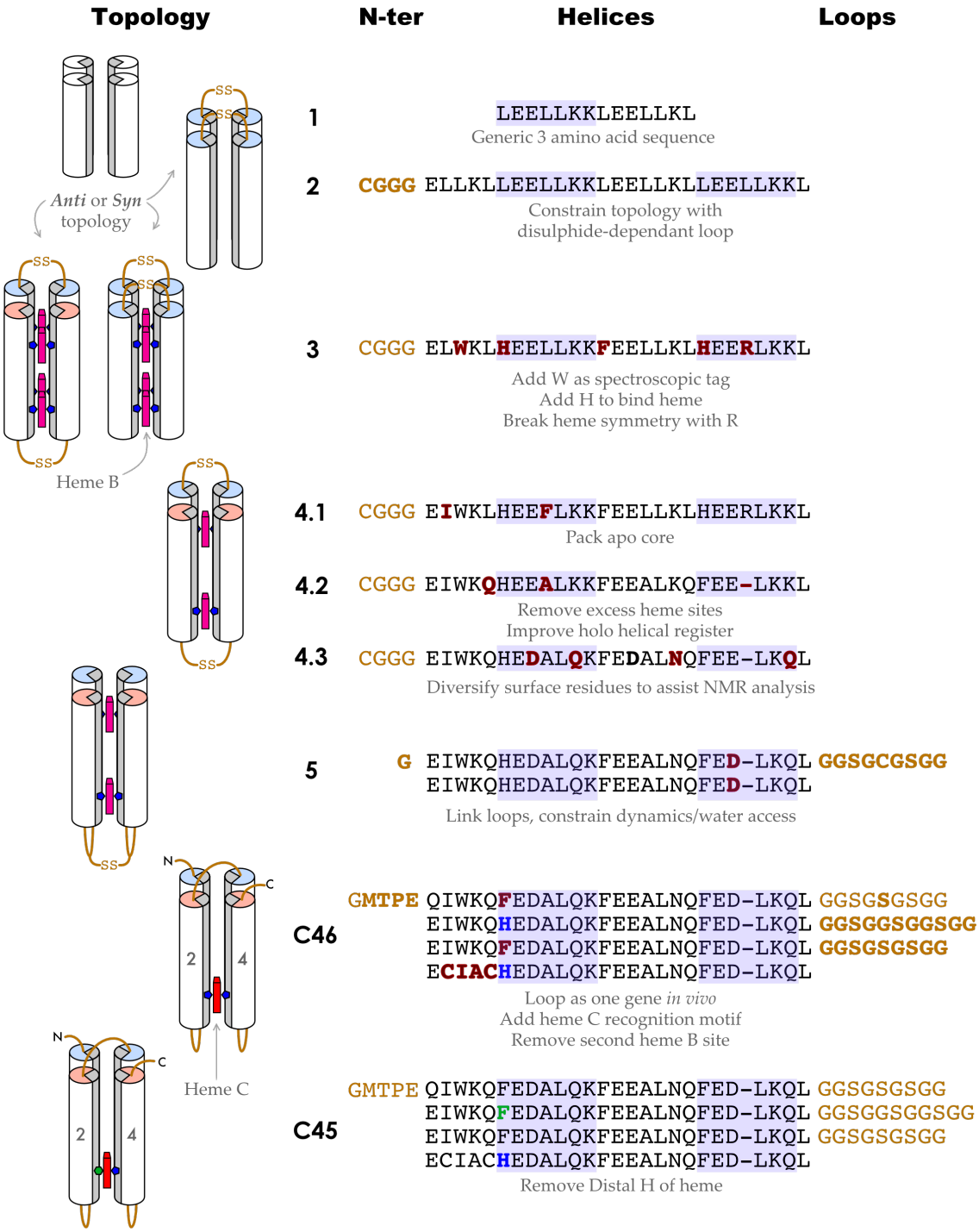


Figure 1.4.1 Maquette Family History.
The shaded region in the helix highlights one heptad. Loop sequences are displayed in yellow. The red bold letters in the sequence have been altered since the previous iteration.

A **third** generation incorporated heme B by mutating two of the core leucine residues to histidine (at position 10 and 24). Each pair of histidine residues were able to ligate the heme iron above and below the plane of the porphyrin ring, sequestering four heme cofactors to the interior of the protein [63]. In addition to this primary mutation, a leucine position

between the heme was changed to phenylalanine since they are commonly found near hemes in natural proteins. In order to distinguish the heme binding sites from one another an interior leucine was replaced with an arginine.

The **fourth** generation was focused on simplifying and validating the tertiary structure of the maquettes. Two of the four heme binding sites were removed to simplify the maquette. A model of the apo structure showed a >50˚ rotation around the helical axis on removal of heme from the ligating histidines. Changes were made to the residues that had moved, making them more amenable to residing at the hydrophobic/hydrophilic interface. To restrict the conformational flexibility at these positions β-branched aromatic residues were introduced. These changes as well as the stabilising hydrogen bond between histidines in the apo form allowed the tertiary structure of the bundle to be seen by NMR and X-ray crystallography (Figure 1.4.1-4.1)[64, 65]. A deletion was introduced to better align the binary patterning of the core and results in a helix that is 26 residues long [66].

1.4.1 Oxygen binding maquette

This generation contains glutamate at positions 11, 18 and 25 in order to try and weaken the histidine-iron ligation at the distal side of the heme. However, this helical strain was not enough to allow oxygen to bind stably at room temperature. Binding CO and illuminating at -15 ˚C showed that exchange to O_2 was possible with 100% yield. This unstable binding of O_2 at room temperature was thought to be a result of easy access to the heme face allowing protons and water to facilitate electron transfer to from ferrous heme to oxygen, a trend seen in natural globins [67]. The **fifth** generation of heme maquettes, focused on O_2 binding, attempted to constrain the motion of the maquette to exclude water from the heme. The NMR result of maquette 4.3 (Figure 1.4.1) showed that the inter-helical interface of the third heptad where the second heme binding had been removed was the most motile section of the helix [62]. To remedy this, the helices were looped in together at the most mobile interface. Disulphide-mediated dimerization of the loops results in a full 4HB in a monomeric 'candelabra' topology [62]. NMR Deuterium exchange experiments showed that the exchange of core backbone residues had improved from around 15 min (Figure 1.4.1-4.2) to several hours (Figure 1.4.1-5). Crucially, O_2 was now able to bind stably at room temperatures with exchange coefficients comparable to natural globins.

1.4.2 *c*-type heme maquettes

The **sixth** generation of maquettes were first modified by looping all four helices together. Two helices are linked by two short loops and one long loop to facilitate the same helical topology as Figure 1.4.1-5. This sequence was transformed into a plasmid with an N-terminal signal sequence to target it for periplasmic export, a hexahistidine tag for purification and a Tobacco Etch Virus protease (TEV) cleavage site to remove the hexahistidine tag (shown in Figure 1.4.3). The maquettes could now be expressed *in vivo* and purified. The next step was to further simplify the maquette by removing one of the heme B binding sites and replacing the other with a Heme C recognition sequence (CxxCH). This sequence is recognised by the endogenous Cytochrome *c* biogenesis system I (Ccm) in *Escherichia coli* (*E. coli*). The Ccm covalently attaches heme B to the cysteines of the recognition sequence with the histidine becoming the proximal ligand to the heme iron and the remaining histidine in the 4HB ligating the distal face of the heme. Proteins with a covalently bound heme B are then termed *c*-type cytochromes. The resulting maquette (Figure 1.4.1-C46) is a 4HB with a single *c*-type heme and is a *c*-type heme maquette (CHM).

A covalently bound heme offers several advantages. The holo-protein can be full expressed *in vivo*, unlike the *in vivo* expressed heme B maquettes which needed to bind heme after purification *in vitro*. This has profound implications for developing the maquettes with techniques like directed evolution, or for use in synthetic pathways *in vivo*. The recognition sequence for *c*-type heme (CxxCH) makes three contacts with the heme. This directional binding is asymmetrical and constrains the orientation the heme can face. The most common rotamer for *c*-type heme proteins (m166) has the heme rotated 45˚ from perpendicular to the helix it is attached to [68]. From a design perspective incorporating a *c*-type heme allows the designer to control the position and orientation of the heme. Work preceding the design of C46 (in Figure 1.4.1) showed that the orientation of the heme propionates around the bundle significantly altered the stability and heme incorporation [69]. Orientations with the propionates facing into the bundle were less stable. Having the CxxCH motif on the forth helix was seen to be the most stable allowing the propionates to face the solvent rather than the core [69]. C46 is named so because it is a CHM, with the *c*-type binding motif on the 4th helix, and the heme iron is bis-histidine ligated resulting in all 6 of the octahedral ligand sites being filled by the porphyrin ring or protein.

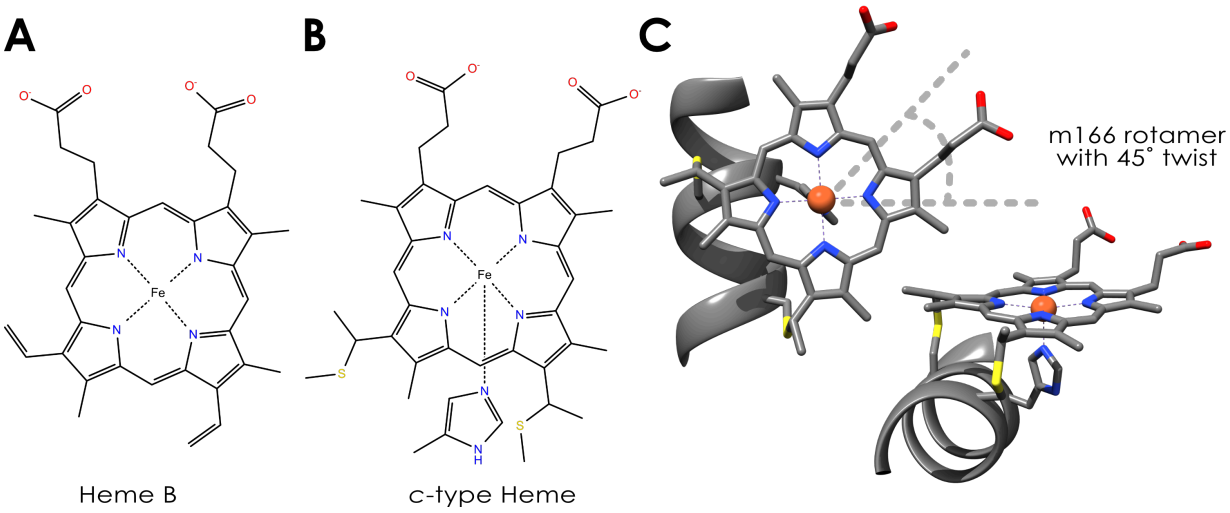


Figure 1.4.2 Heme B and C comparison. ***(A)*** *shows a skeletal representation of heme B.* ***(B)*** *shows a skeletal representation of c-type heme. The sulfurs and imidazole ring belong to cysteine and histidine protein side chains truncated at the Cβ carbon. (C) shows the structure of the most common c-type heme rotamer found on a-helical structure, m166 [68].*

1.4.3 A penta-coordinate heme maquette

The **seventh** generation of maquettes in this family tree was aimed at producing oxygenase activity. As discussed further in Section 1.7, natural heme-containing peroxidases and oxygenases generally have a penta-coordinate geometry around the heme iron leaving the distal side of the heme free to bind water in the resting state. Taking inspiration from these natural enzymes a single point mutation was made to remove the distal ligating histidine. This was only possible using a covalently bound c-type heme maquettes as removing the histidine from a heme B binding maquette eradicates stable binding. This penta-coordinate CHM, named C45, exhibits peroxidase activity and is the central subject of this thesis [70].

To complement the maquette design methodology this project utilizes computational modelling. Advances in computing power, quantum mechanical (QM) approaches and experimentally validated force fields makes computational modelling often the most useful and effective method of probing reactivity and protein dynamics. This project will utilise different computational methods to study structural dynamics and stability and possible binding of ligands. The inclusion of computational modelling to inform the design produces a workflow seen in Figure 1.4.3. The current CHM model is not directly based on an experimental structure, as there is not one available. The CHM model was made using a related maquette protein, crystallised in 2003 (Figure 1.4.1-4.2) [65].

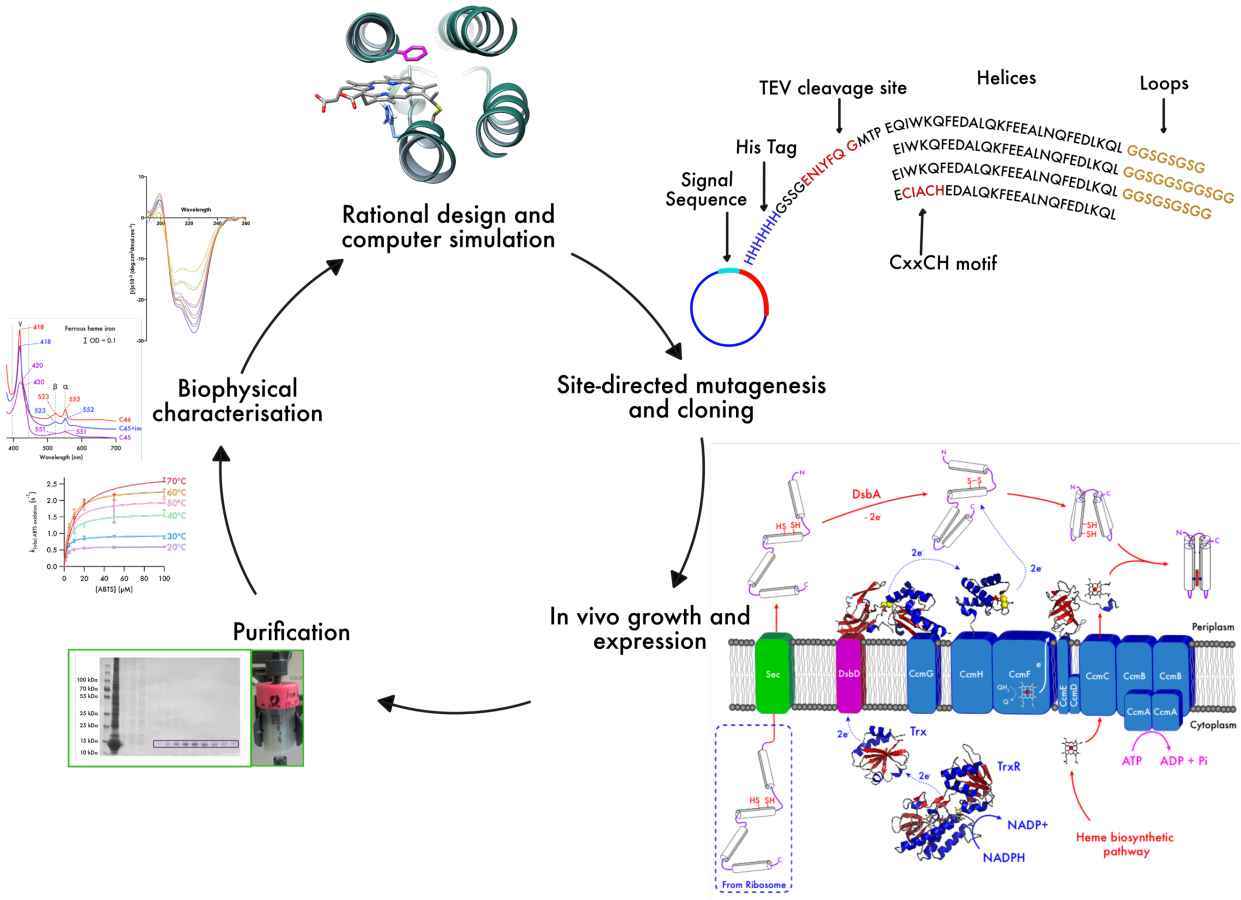


Figure 1.4.3 Maquette design cycle.

Starting with rational design aided by computational simulation and moving clockwise around the cycle gives the maquettes design cycle. The image at "In vivo growth and expression" is a schematic of heme insertion by the Ccm. It is taken from Watkins et al 2012 [71].

1.4.3.1 Flexibility

Like some natural proteins, heme binding in maquettes generally seems to increase protein rigidity. Cytochrome b_{562} is a 4HB thought to be involved in the bacterial electron transport chain [72]. It is considered molten globular in the Apo form [50]. A molecular dynamics study suggested that while NMR assignment was hindered due to high motility of the amino acid side chains, the overall protein structure was still highly ordered [50]. Binding heme B constrains the protein into its native structure. Equally, Apo-cytochrome *c* is considered molten globular and heme binding is required to attain the native structure [73]. Hp-1 is a homodimeric di-heme binding maquette shown in Figure 1.4.1 labelled 4.3. On binding heme 1D ^{1}H NMR spectrum sharpen as the protein transitions from a molten globular structure to a more native-like conformation.

1.4.3.2 Redox properties

The electrochemical properties of the heme in the maquettes can be adjusted by changing the environment around the heme. The electronic midpoint potential (E_m) of most maquettes is negative [58]. The range in E_m of heme-containing maquettes is between -290 and 6 mV [58, 74]. Increasing the external charge of a maquette (by +29) has been seen to increase the E_m by 100 mV [75]. Adding positive or negative residues near the heme has also been shown to have a more subtle effect in E_m [75]. Figure 1.5.2 has more information about the redox properties of different proteins to put this information in context.

1.5 Cofactors driving redox chemistry

The functional groups of amino acid side chains are well suited to acid–base reactions, the formation of certain types of covalent bonds, and charge–charge interactions, but are less well equipped for the electronic requirements of redox chemistry. Consequently, redox enzymes tend to incorporate cofactors to impart function. Cofactors are non-proteinaceous molecules required by some enzymes for function. These can be metal ions (commonly iron, copper and zinc cations), metals within specialised coordination complex's (hemes, and chlorophylls) or small organic molecules (such as NAD or flavins). Oxidoreductases are enzymes that catalyse the transfer of electrons from electron donor (reactant) to an electron acceptor (oxidant). There are many different types of Oxidoreductase; dehydrogenases, oxygenases, peroxidases and hydroxylases are some examples. Figure 1.5.1 gives some examples of different redox cofactors. Coenzymes are cofactors that are not bound for the whole reactive cycle of the enzyme. Nicotinamide adenine dinucleotide (NAD), which is used in metabolic cycles in all domains of life, is usually oxidised and replaced before the next catalytic cycle. When a cofactor is tightly associated with an enzyme, like the heme groups in cytochrome P450's or the c-type cytochromes, it is known as a prosthetic group [12].

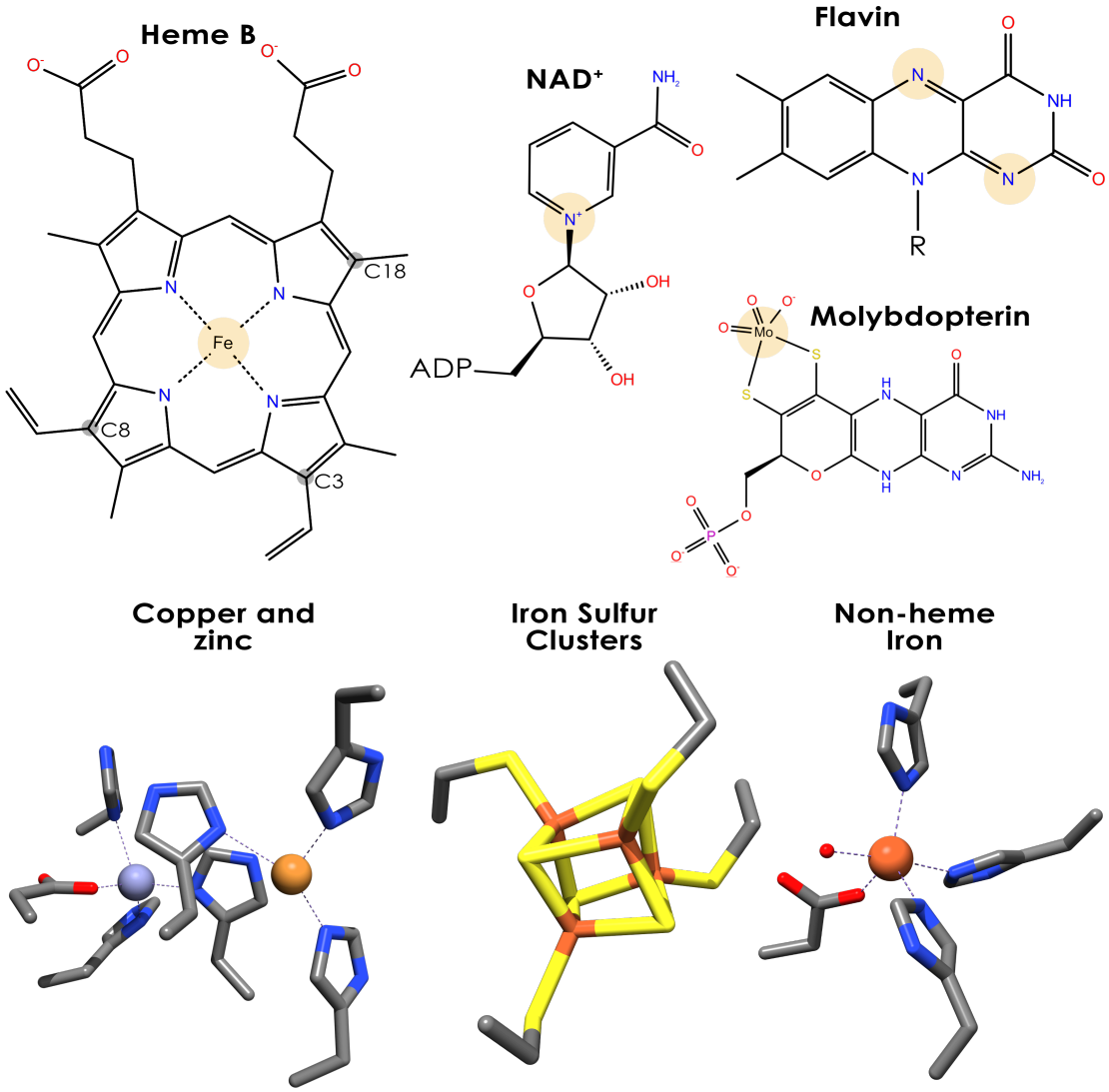


Figure 1.5.1 Redox cofactors. *The redox cofactor represented in skeletal form have the redox-active part of the cofactor highlighted in orange. Heme B has 3 carbons of the porphyrin labelled that are the site of modification for Hemes A O and C (see Figure 1.4.2). Nicotinamide adenine dinucleotide (NAD) is shown in its oxidised form. Flavin group is shown in its reduced form. The R-group on the Flavin can represent either adenine mononucleotide (FMN) or adenine dinucleotide (FAD). Molybdopterin is a nitrate reductase associated cofactor [76]. The last row show redox active metal ion with their ligating protein side chains. The Copper and zinc protein shown is superoxide dismutase [77]. The 4Fe-4S cluster is the active site ferredoxin from Bacillus thermoproteolyticus [78]. The non-heme iron example a different superoxide dismutase from Sulfolobus acidocaldarius [79].*

1.5.1 Electron transfer

Generally, the distance between redox centres is calculated by the edge-to-edge distance of the conjugated system. The rate of electron transfer (k_{et}) is exponentially related to the distance between the two redox centres (R). The Moser-Dutton equation (shown below) builds on the work of Marcus, Fermi and Hopfield, repurposing Marcus theory for electron

transfer for use within an insulating protein environment [80–82]. The driving force for electron transfer is the change in Gibbs free energy (ΔG in eV) and the reorganisation energy (λ). In this equation the packing density of the protein environment (vacuum-like or bond-like) has been substituted for an average value in a normal protein (ρ = 0.76) and the remaining coefficients are tailored for exergonic electron transfer reactions at room temperature.

$$logk_{et} = 13 - (1.2 - 0.8\rho)(R - 3.6) - 3.1(\Delta G + \lambda)^2/\lambda \quad (1.1)$$

Charge transfer from one redox centre to another changes the environment around each centre and typically requires the reorganisation of atoms near the redox centres, often the solvent. The energy associated with atomic movements to support the new electronic environment is the reorganisation energy (λ). In a polar environment the response to electronic changes is larger with a λ around 1-1.4 eV. In a non-polar environment the λ is reduced, taking a value between 0.2-0.7 eV [83]. An electron transfer to cofactors with a high charge density incurs a greater λ than to a cofactor with a large conjugated network and a low charge density. The exponential relationship between distance (R) and rate of transfer (k_{et}) means that as the distance increase to 14 Å electron transfer reactions become so slow they are biologically irrelevant. Redox partners in nature are rarely found more than 12 Å apart [84].

The **midpoint potential (E_m)** of a redox centre is an experimentally derived value. It is quoted in Volts versus the Nernst hydrogen electrode (NHE), where an equimolar mixture of the reduced and oxidised states are found. The midpoint potential is related to the ΔG of electron transfer by the equation below. Where n is the number of electrons and F is the Faraday constant. As the difference in midpoint potential (ΔE_m) between two redox centres increases so does ΔG.

$$\Delta G = -nF.\Delta E_m \quad (1.2)$$

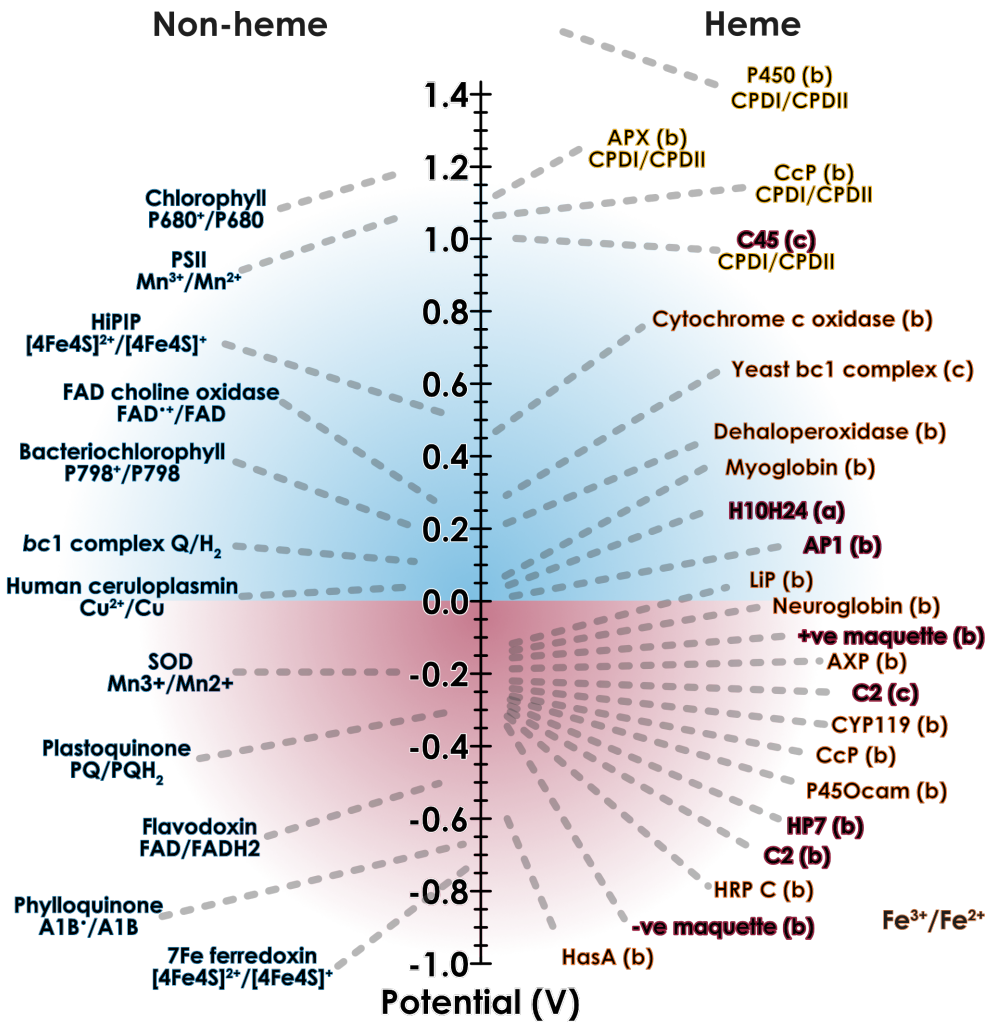


Figure 1.5.2 Redox midpoint potentials compared. *The potentials for* ***non-heme*** *redox cofactors shown are: 7Fe ferredoxin [85], phylloquinone [86], flavodoxin [87], plastoquinone [88], SOD [89], human ceruloplasmin [90], bc1 complex [91], bacteriochlorophyll [92], FAD choline oxidase [93], HiPIP [94], photosystem II [95], and chlorophyll [96]. The potentials for* ***ferrous to ferric*** *heme is show for the following proteins: HasA [97], horseradish peroxidase C [98], cytochrome P450cam [99], Cytochrome c peroxidase (CcP) [100], CYP119 [99], Ascorbate peroxidase [101], neuroglobin [102], Lignin peroxidase (LiP) [103], myoglobin [104], dehaloperoxidase [105], yeast bc1 complex [106] and cytochrome c oxidase [107]. The potential for* ***compound I to compound II*** *(CPDI/CPDII) is shown highlighted in yellow for: C45 (maquette used this project), CcP [108], AXP [101],and P450's [109]. The maquette redox potentials are highlighted in blue. -ve and +ve maquette are maquettes with high net surface charges [58]. HP7 Heme B maquette (figure 1.3.1.5) [110]. C2 is a heme B and C binding protein (both redox potentials are shown) [111]. AP1 is a lipophilic maquette [74]. H10H24 is a heme A binding maquette [54].*

1.5.2 Heme

Porphyrins are a class of cofactors used by a wide range of oxidoreductases, from mitochondrial proteins that take part in oxidative phosphorylation (complexes II, III and IV) to the reaction centre at the core of the photosynthetic apparatus. At the heart of all porphyrins is a tetrapyrrole conjugated system. Porphine is the simplest macrocyclic

porphyrin. Modifications of the porphine macrocycle, periphery and the metal ion, give rise to: hemes, chlorins (like chlorophyll a) and corrins (like cobalamin). These modifications explain the huge diversity in chemical properties of different natural porphyrins. The pyrrole nitrogens at the centre of the porphyrin can coordinate a number of metal ions, occupying the equatorial coordination sites of the metal. Different metals exhibit different the geometries around the axial ligands of the porphyrin. Palladium, zinc and iron porphyrins will present square planar, square pyramidal and octahedral geometries respectively [112].

Nature can fine-tune the properties of porphyrins by changing the protein environment around the cofactor. The axial ligating protein residues generally have the greatest effect on the electronic properties but changes in the second shell around the cofactor can also produce a marked effect. Cysteine, histidine, tyrosine and methionine are all common axial ligands for oxygenase, peroxidase, catalases and electron transport proteins like Cytochrome *c* [102, 113–115]. Natural iron protoporphyrins are known as hemes and all have two propionic acid groups attached at one edge of the macrocycle. At physiological pH's these acid groups are deprotonated with an approximate pK_a of 3. These charged groups are generally orientated towards the solvent in heme proteins with the macrocycle buried in the hydrophobic core of the protein. The electrostatic contacts made with the porphyrin propionates can also modulate the electronic properties of the heme as well as confer a better binding affinity of the heme [116]. The nomenclature around heme sees the letter of the heme capitalised when referred to outside of a protein, but is italicised in lowercase when as a prosthetic group, eg heme B inside the main subunit of complex III is known as cytochrome *b*.

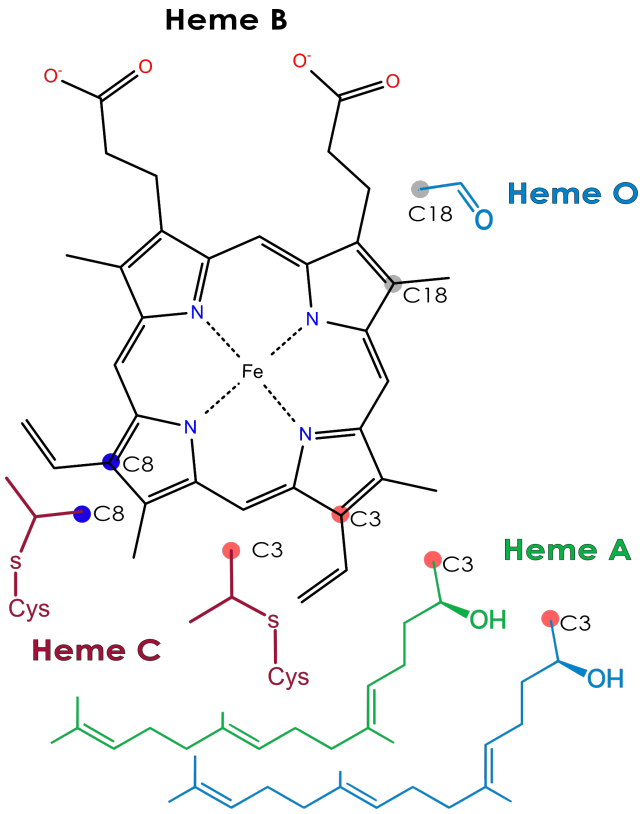


Figure 1.5.3 Heme modifications. *Heme B is shown in full. Heme A has the green modification to carbon C3. Heme O has both blue modifications to C3 and C18. Heme C has the red modifications shown for C3 and C8*

Heme A, B, C and O are all biologically common hemes that differ by the organic groups attached to sides of the macrocycle (not the side containing the propionates). Heme O and A differ only by the attachment of a methyl instead of a formyl group one carbon along from the second propionate attachment. This change confers a 180 mV decrease in midpoint potential between Heme O and A [116]. Heme C is a misnomer as it refers to heme B that has been covalently attached to a protein. Proteins with this modification are known as *c*-type cytochromes. The covalent thioester linkages are attached via the vinyl groups of heme B and only confer a modest change in redox properties [69]. As hemes are generally buried in the hydrophobic core of proteins it is common to see π-stacking of the heme with nearby phenylalanine residues. Heme B binding in myoglobin is dependent on π-stacking interactions between phenylalanine and the heme. Mutating phenylalanine 46 results in a greater than 10-fold loss in heme binding [117]. In H-NOX (Heme Nitric oxide and/or Oxygen binding)

domains π-stacking has been seen to decrease the midpoint potential of the heme by 40 mV and increase the binding affinity of O_2, CO and NO to the heme iron [118].

In mammals, the most abundant porphyrin is heme B (iron protoporphyrin IX). It is used to transport electrons around the respiratory transport chain, in haemoglobin and myoglobin to transport and store oxygen, and also has an enzymatic function [119]. Its role in enzymatic activity predates the understanding that the observed activity came from a heme containing enzyme. HRP was discovered in 1903 and CcP in 1940 [120, 121]. Owing to their relative ease of purification and HRP's broad substrate reactivity they represent some of the best characterised enzymes. Heme containing oxygenases, liver cytochrome P450s, were first discovered in 1958 and later classified as heme proteins in 1964 [122–124]. The cytochrome P450 superfamily are of particular interest to the pharmaceutical industry due to their key roles in drug and xenobiotic metabolism. Natural oxygenases are able to break up molecular oxygen and insert it into an unreactive organic molecule in a tightly controlled reaction [125]. Globins in the resting state are typically in the ferrous (Fe^{2+}) oxidation state but cytochromes P450 are in the ferric (Fe^{3+}) oxidation state at rest and access reactive ferryl (Fe^{4+}) intermediates during their reactive cycle.

1.6 Heme oxidoreductases

Heme peroxidases and oxygenases are both oxidoreductases. Peroxidases typically use peroxide (H_2O_2) to oxidise two equivalents of electron-donating substrate but do not usually catalyse oxygen-transfer reactions. Oxygenases use molecular oxygen (O_2) as their oxidant, inserting one of the oxygen atoms into the substrate. Oxidation with O_2 of most biological molecules is thermodynamically favourable. However, these oxidations do not occur as O_2 is a paramagnetic molecule and biological molecules are spin paired making it a spin forbidden process. This places large kinetic barriers on these reactions forcing nature to incorporate transition metal cofactors to circumvent it [126].

1.6.1 P450's

P450's monoxygenases come from the P450 superfamily of cysteine ligated heme B containing proteins. They are known as monoxygenases as only one of the oxygen atoms of molecular oxygen is used to oxygenate the other being reduced to water. P450's were first shown to be enzymatically and physiologically important in 1963 where the 450 nm pigment protein in adrenal cortex microsomes was seen to catalyse C-21 hydroxylation of 17-hydroxyprogesterone [127]. Since then P450's have been seen to catalyse a diverse set of reactions on both endogenous physiological substrates but also drugs and chemicals foreign to the organism (xenobiotic). Almost all eukaryotic P450s are membrane associated, residing in the inner mitochondrial membrane or the endoplasmic reticulum. Over 450 different P450's have been discovered and can be found in virtually all mammalian tissues. Liver microsomal P450s in mammals mediate membrane detoxification by oxygenating and partially solubilising toxic hydrophobic xenobiotics, allowing them to be excreted. P450 are blind to the knowledge that a particular xenobiotic may be a therapeutic molecule and can also convert these helpful compounds to toxic metabolites. Therefore, drug metabolism *via* cytochromes P450 has become a significant consideration for the pharmaceutical industry. Important for predicting drug interactions, toxicity and pharmacological lifetime [128, 129].

P450's are also of interest in biotechnology and synthetic biology as the ability to functionalise unactivated carbon-hydrogen bonds in a regio- and stereospecific manner is a challenging process. Natural P450's catalyse a variety of interesting and intrinsically valuable reactions: hydroxylations, N- or O-dealkylations, epoxidations and Baeyer–Villiger oxidations [130]. These reactions could be exploited to provide a more environmentally friendly alternative to traditional chemical synthesis. P450's have also been engineered to perform non-biological carbene and nitrene transfer reactions with superb stereoselectivity [131]. P450's use a conserved cysteine residue for the proximal axial heme B ligand. Mutating cysteine to histidine results in an improved cyclopropanation activity in the synthesis of levomilnacipran [132]. P450's designed and optimised by directed evolution can now be

bought as commercial products. Arnold *et al.* currently hold 57 patents for various P450's showing a growing commercial appreciation for these enzymes [133].

1.6.2 Intermediates

The catalytic cycle of both heme-containing oxygenases and peroxidases are very similar and access the same highly redox active intermediates. These similarities have allowed insights to be gained from the study of one heme oxidoreductases family to be transferred and compared to the others [119]. As seen in Figure 1.6.1, generally in the P450 superfamily, the water bound at the distal face of the heme is displaced by the binding of substrate. Increasing the midpoint potential (E_m) of the ferric (Fe^{3+}) heme by about 100 mV [134]. This change in electronic potential triggers the reduction of the ferric heme (Fe^{3+} to Fe^{2+}) by NADPH. This reduction is facilitated reduction by cytochrome P450 reductase. These organic substrates tend to bind orientating the target C-H bond in close proximity to the heme [135]. P450cam was the first P450 to have its structure determined with its natural substrate camphor in the active site [136]. The structure showed that the reactive carbon of camphor is just 3.4 Å away from the heme [137]. Molecular oxygen can then bind and the complex is reduced a second time from an NADPH-derived electron passed from cytochrome P450 reductase. This peroxyferric intermediate is known as compound 0. All heme-containing oxygenases progress through the peroxiferric species in their reactive cycle, and it is after this point that the difficult and interesting chemistry of oxygen bond cleavage takes place. If peroxide is bound instead of O_2 P450s follow a mechanism known as the peroxide-shunt pathway that proceed similar to the peroxidase reactive cycle [138]. The activated oxygen, in the form of peroxide, is used to make compound 0 directly.

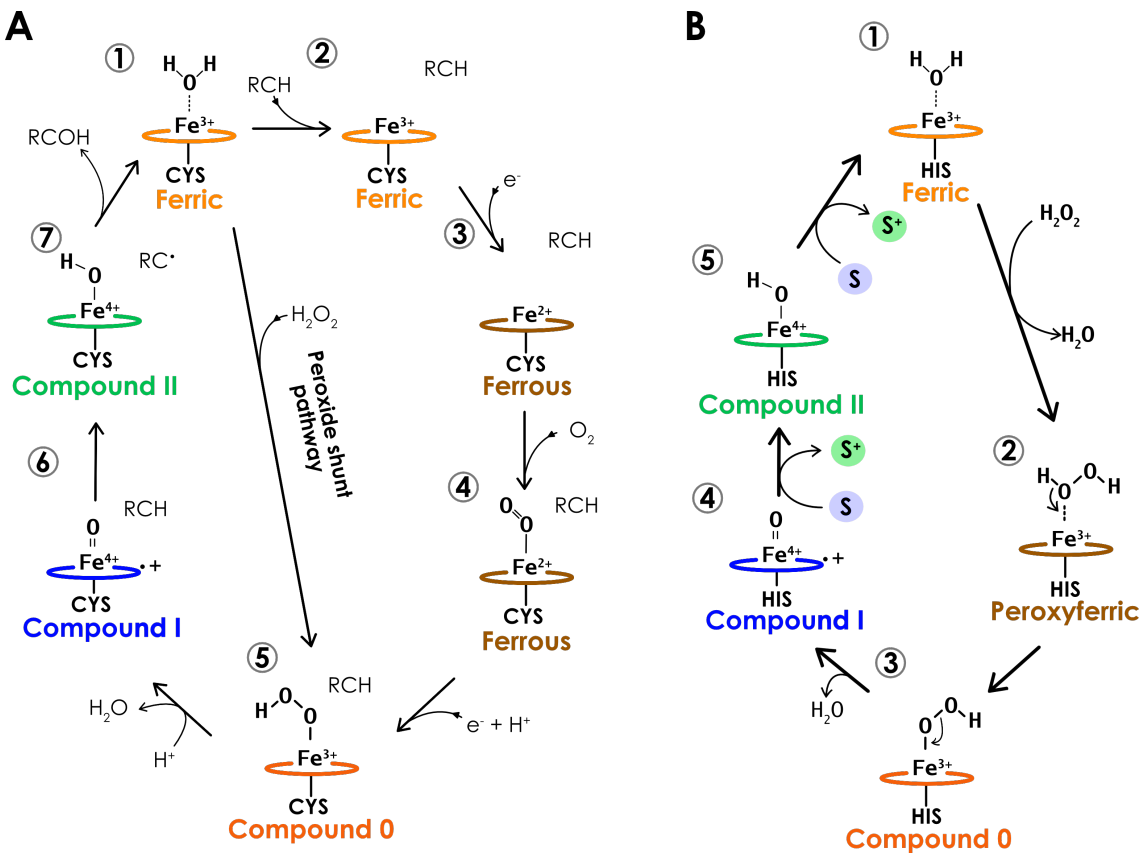


Figure 1.6.1 Reaction cycle of P450's and peroxidases. ***(A)*** *shows the reactive cycle of a typical P450. 1. The resting state Cys-Fe*$^{3+}$*has water bond at the heme iron. 2. Substrate binding in the distal cavity displaces the water ligating the heme iron and produces a 100 mV increase in the heme midpoint potential (*E_M*). 3. In humans this triggers the reduction to the ferrous iron oxidation state with NADPH mediated by cytochrome P450 reductase. 4. After binding to the heme, molecular oxygen is reduced by a second equivalent of NADPH, mediated by cytochrome P450 reductase. 5. Compound 0 is formed by the reduction and subsequent protonation of the bound oxygen. 6.Compound 0 is then protonated by a protein side chain in the active site and condenses. This produces the compound I porphyrin radical cation. 7. Compound I oxidises the reactive group on the substrate producing compound II and a carbon radical cation. The carbon radical then immediately rebounds abstracting the hydroxyl group from the heme to form an alcohol* [139]*. Product exiting the active site returns the heme to the resting state. The Peroxide shunt pathways shows how the use of an activated oxygen, in the form of peroxide, is used to make compound 0 directly [138].* ***(B)*** *shows a typical peroxidase reactive cycle. 1. Peroxidase have a ferric resting state with water bound. 2. Peroxide binding produces a peroxyferric complex that is then deprotonated. 3. This deprotonation forms compound 0. 4. Compound 0 is protonated, most likely by a water mediated mechanism involving the distal active site residues, and compound I is formed. Compound I is generally a porphyrin cation radical but protein radical cations have been observed in some peroxidases (see Section 1.8.4). 5. Compound I is reduced by an electron donating substrate to form compound II which is then reduce by a second electron donating substrate. The reductions are coupled to protonation of the ferryl oxygen in both cases.*

The P450 monogeneses strategy reductively activating molecular oxygen requires at least stoichiometric quantities of a reducing substrate [140]. Generally, compound I is regenerated using nicotinamide cofactors. Not only are these cofactors expensive to produce many

reducing equivalent are lost unproductive side reactions [141]. These side reactions most commonly feature O_2 as an electron sink and subsequently produce reactive oxygen species (ROS). The peroxide-shunt pathway removes the problem of using reducing cofactors [138]. As a result, there is considerable interest in engineering P450 to become better peroxygenases as well as the prospect of using fungal heme-thiolate peroxidases which naturally exhibit peroxygenase [142, 143].

The O=O bond could be broken homolytically, generating two hydroxyl radicals, which would likely cause problems in any enzyme active site. Heterolytic cleavage is therefore selected for in nature. The products of lysis in this case effectively are a bare oxygen atom with only 6 valance electrons and water. The oxygen bound to the iron oxidises the heme producing a porphyrin radical cation intermediate known as compound I (Cys-Fe^{4+}=O:Porphyrin$^{\bullet+}$). Compound I is then able to abstract a hydrogen atom from the substrate C-H bond, reducing the compound I radical to compound II (Cys-Fe^{4+}-OH) and producing a carbon radical on the substrate. Finally Compound II is reduced back to the resting state (Cys-Fe^{3+}) as the hydroxyl group is transferred to the substrate producing a hydroxylated carbon [144]. Oxygen-activating maquettes could one day be included in a biocompatible system, offering new functionality in pathways to metabolise drug molecules.

1.7 Heme peroxidases

As previously mentioned, peroxidases were the first heme enzymes to be discovered. The X-ray structure of CcP was added to the protein data bank (PDB) in 1980 when there were only 57 other structures archived [145]. **Myeloperoxidase,** lignin peroxidase (LiP), ascorbate peroxidase (APX) and HRP followed [146–149]. At the time of writing (September 2018) there are now 705 heme peroxidase structures from over 200 different organisms [9]. In general, peroxidases catalyse the reduction of hydrogen peroxide to water coupled with two single electron oxidations of an electron donating substrate.

$$HOOH + 2e^- + 2H^+ \xrightarrow{peroxidase} 2H_2O \qquad (1.3)$$

Peroxidases are grouped in to three main classes. Class I are intracellular peroxidases such as CcP. Class II and III are extracellular plant and fungal peroxidases respectively. Heme peroxidases have a relatively well-conserved active site structure around the heme. On the

proximal face of the heme they are ligated by a histidine residue (pHIS) that forms a hydrogen bond with an aspartic acid residue (pASP), also on the proximal side of the heme. In CcP and AXP a proximal tryptophan residue (pTRP) also hydrogen bonds with the pASP and lies parallel to the dHIS very close to the heme edge. In CcP it is this pTRP that forms the protein radical cation compound I. Other peroxidases contain a phenylalanine residue in the place of the pTRP found in CcP and AXP.

Peroxidases also have a conserved histidine (dHIS) and arginine (dARG) residue in the distal heme pocket. AXP and CcP also have another tryptophan on at the distal face of the heme (dTRP) but this position is normally filled by a phenylalanine in other peroxidases. The dARG (and dTRP in CcP and AXP) have been proposed to help stabilise the formation of compound I. dARG is seen in two different conformations, “in” and “out”. In the resting state it is “out” but in compound I structures it is “in”, close to the ferryl oxygen. The most kinetically important residue of the distal site is the dHIS. It is proposed to be involved in proton transfer at various stages of the catalytic cycle which will be discussed later in the chapter.

All classical peroxidases exhibit a pentacoordinate geometry around the heme iron. Leading to water occupying the final ligation site on the distal heme face in the resting state. In the resting state the iron is in the Fe^{3+} oxidation state and is subtly bent out of plane towards the proximal ligand with respect to the heme. Like all classical peroxidases, HRP in the resting state exists as a mixture of spin states S=3/2 (low spin) and S=5/2 (high spin) [150].

1.7.1 Substrate interactions

Peroxidases will generally turn over a broad range of compounds, with phenolic substrates being the most commonly studied. HRP exemplifies the trend of substrate promiscuity in peroxidases. It has been characterised with many different electron donating substrates but its native substrate is still unknown. Differences between the reactivity of the ferryl heme (compound I) and interactions with substrate control the diversity of functionality across the

heme peroxidase family [151]. Phenols and aromatic amines are the most common small molecule-reducing substrates. Hydrogen peroxide can also be replaced with peroxy acids with the general formula ROOH [152].

1992 saw the first crystal structure of CcP and Cytochrome *c* in complex and was the first structural characterisation of a peroxidase with its natural substrate [153]. This milestone added fuel to the fire of rigorous debate around the binding and inter-protein electron transfer between Cyt *c* and CcP [154]. Unlike HRP's green compound I intermediate, CcP produces a red intermediate that has a characteristic g ≈ 2.0 EPR signal [155]. The signal has been identified as a tryptophan radical (TRP191$^{\bullet+}$) [156]. Electron transfer proceeds from the tryptophan radical along only a few peptide bonds to Cyt *c*. There is also on-going debate around the two binding sites identified for Cyt *c* on CcP, and to what extent they are physiologically relevant [154]. Early studies of CcP provided extremely important information in understanding interactions of protein-protein electron transfer [154]. However, it was less useful for understanding the nature of binding interactions with small molecule ligands.

NMR studies with HRP and those using phenylhydrazine to mimic small molecule substrates indicated that substrates were interacting with the ∂-heme edge [157–160]. In this regime it was assumed that small molecules bind close to the heme and electron delivery occurs *via* the ∂-heme edge, with the CcP's protein-protein transfer seen as an outlier. First the crystal structure of MnP and then ascorbate peroxidase AXP challenged this general mechanism [161, 162]. Both structures show the substrate bound on the γ-heme edge with the heme propionates hydrogen bonding directly with the substrate. This evidence of an alternative election donation pathway through the heme propionates highlights the natural promiscuity of peroxidases.

Most peroxidases can use small phenolic substrates as a source of electrons. Guaiacol is among the most studied as it is readily converted into its green-coloured tetramer, tetraguaiacol. In 2012 a crystal structure of CcP co-crystallised with guaiacol located two

possible binding sides [163]. One is a hydrophobic pocket around 23 Å from the heme, too far for direct electron transfer. EPR data suggests it could be catalytically relevant if transfer proceeded through a Tyr71 radical which is between the heme and the hydrophobic pocket [164]. This type of long-range electron transfer has also been suggested to take place in versatile peroxidase (VP) and dye-decolorizing peroxidase (DypB) [165, 166]. The second is around 15 Å from the heme edge but is not thought to be catalytically active. Introducing a mutation that eradicates binding at this site does not affect the steady-state turnover. To add more uncertainty, inhibition studies with isoniazid, which is known to bind at the heme edge of CcP indicates that both compounds bind at the same site [163].

1.7.2 Lignin peroxidase

LiP is a class II peroxidase that is produced by wood-degrading fungi (white wood-rot fungus *Phanerochaete chrysosporium*) [167]. LiP has a reactive cycle similar to most peroxidase. LiP is secreted from fungi along with Veratrole alcohol (VA, 3,4-dimethoxybenzyl alcohol), H_2O_2 generating enzymes and manganese dependent peroxidase (MgP) and VP during the process of lignin degradation [168]. LiP and VP are able to oxidise high redox potential substrates like VA and other lignin breakdown products that HRP and other lower potential peroxidases cannot [169]. HRP will oxidise polycyclic aromatic compounds only with ionization potentials below 7.35 eV were as LiP oxidises polycyclic aromatics with ionizing potentials of approx. 7.5 eV or lower [170, 171].

VA has been proposed act as a diffusible redox mediator that could allow the oxidising power of the enzyme to be made available to degrade lignin, a large and insoluble polymer. In a similar regime to MnP that oxidises Mn^{2+} to Mn^{3+} which is released from the enzyme as a diffusible redox mediator. Alternatively, it has also been proposed that the LIP-$VA^{\bullet+}$ complex is the redox pair with the LiP stabilising the relatively short-lived $VA^{\bullet+}$ radical. VA turnover in the recombinant LiP isoenzyme H8 (LiPH8*) is completely lost on mutation of Trp171 which is about 12 Å from the α-heme edge in an acidic pocket [172]. Reactivity towards other lower potential peroxidase substrates was retained. The current understanding is that LiP has

two possible binding modes: at the heme edge like classical peroxidases; and for high redox potential substrates another mediated through a Trp171 radical in an acidic pocket at the protein surface.

The general picture of peroxidases and their substrates is that for non-specialised substrates there seems to often be multiple-low affinity binding sites. Most evidence suggests that kinetically-relevant binding occurs at the δ-heme edge. For more specialised peroxidases like AXP, MnP and CcP generally have a better-defined binding site for their native substrate and are less reactive with non-native substrates. Long range electron transfer is also not uncommon. It occurs *via* production of protein radicals, and both $Trp^{\bullet+}$ and $Tyr^{\bullet+}$ have been observed in such peroxidases.

1.7.3 DyP-type heme peroxidases

Dye-decolorizing peroxidase (DyP) is a member of the DyP-type peroxidase family produced by *Bjerkandera adusta* [152]. This novel family of heme peroxidases is phylogenetically and structurally unrelated to the classical animal and plant peroxidase super families [173]. DyPs have a widespread distribution across bacteria and fungi suggesting that they come from an ancient evolutionally root [174]. Similar to other heme peroxidases they contain *b*-type heme ligated proximally with a conserved histidine residue [175]. DyP has a mix of α-helices and β-strands forming a βαββαβ secondary structure motif with the two βαβ sections sandwiching the distal heme site [175].

Instead of the distal histidine seen in classical peroxidases, DyP has an active site aspartic acid (Asp171) and arginine (Arg329) that are involved in the formation of compound I. Replacing Asp171 with asparagine abolishes enzymatic activity in DyP reinforcing its involvement in an acid–base catalytic mechanism at low pH [175]. DyPs are able to oxidise standard peroxidase substrates like ABTS and 2,6-dimethoxyphenol but also turn over high-redox potential dyes such as Reactive Black 5 and Reactive Blue 5 [176, 177]. They are unable to oxidise veratryl alcohol into veratraldehyde which seems to be a unique feature of LiP and VP. Some DyPs are able to oxidise Mn(II) and a β-aryl ether lignin model compound, suggesting that they may be able to function as a lignin peroxidase [178].

1.7.4 Dehaloperoxidase

The classical heme peroxidases all have well-conserved folds and active sites. The hemoglobins are also heme B binding, α-helical proteins but have a different fold and lack the conserved active site residues seen in classical peroxidases [179]. Hemoglobin-dehaloperoxidase (DHP) is multifunctional haemoglobin that was discovered from *Amphitrite ornate,* a terebellid polychaete. This "ornate" sea worm shares an ecological niche with unfriendly neighbours intent on poisoning it with toxic halogenated phenols and indoles [180]. DHP possesses both oxygen-transport and peroxidase activity that allows it to detoxify 2,4,6-tribromophenol (TBP) to 2,6-dibromo-1,4-benzoquinone (DBQ). TBP is a native substrate but DHP is also active against 2,4,6-trichlorophenol (TCP) and 2,4,6-triflurophenol (TFP) converting them in to their less toxic 2,6-dihalo-1,4-benzoquinone (DXQ) equivalents.

DHP and other haemoglobins have a high degree of structural similarity. Like other oxygen binding globins, heme B is proximally ligated by a histidine residue with another histidine residing at the distal face of the heme. *A. ornata* has two homologues of DHP, DHP-A and DHP-B. While DHP-B differs from DHP-A by only five amino acids, it is substantially more conformational dynamic [181] and is 4.3-fold more active than DHP-A (DHP-B $k_{cat}/K_{M(H2O2)} = 2.1\times10^{4}$ $M^{-1}s^{-1}$) [105]. For a globin, the activity is the highest yet discovered though it is significantly less than classical peroxidases such as HRP (HRP $k_{cat}/K_{M(H2O2)} = 4.6\times10^{6}$ $M^{-1}s^{-1}$) [182].

The conformational flexibility of DHP-B means that the distal ligating histidine is able to occupy two different rotamers. One is inactive and oriented out into solvent, and the other active rotamer interacts with the heme oxygen species. It has been postulated that the binding of substrate to the entrance of the distal heme cavity forces the histidine into the active conformation [183]. Crystallographic binding investigations using TCP and TBP suggest that there are at least two binding sites in DHP. One internal site close to the heme at the α-heme edge with the phenolic OH moiety of TCP able to form a hydrogen bond to the water molecule above the heme iron [184]. The second site, referred to as the external or δ-site, is at the entrance to the heme cavity at the δ-heme edge [185]. There is still debate as to the exact functional nature of the two binding sites. One interesting explanation for the internal site, which may not be in the right position for turnover, is that it may serve to regulate the function of the globing between oxygen binding and dehaloperoxidase activity [184].

DHPs lack the conserved peroxidase proximal residue pASP which is thought to relay charge back in to the ligating pHIS, then to support the delivery of more electron density onto the heme iron. The backbone carbonyl Lys38 in DHP forms a strong hydrogen bond with the pHIS but is not able to relay charge at the same level as a pASP acid group. DHPs have been observed to form tyrosine compound I radicals. Tyrosines 34 and 38 are 5.5 Å and 7.7 Å from the heme edge [186]. Tyrosine 34 is thought to account for the majority of the radical signal [183]. The mechanistic evidence for DHP strongly suggests the reaction proceeds similar to classical peroxidase in two discrete single electron transfers. As a multifunctional protein, the details are even less clear than for other peroxidases. There is evidence to suggest that DHP can initiate peroxidase activity from the ferrous oxidation state with oxygen bound. DHP has an abnormally high reduction potential (+205 mV) compared to other peroxidases [187], which is thought to allow for reduction by DBQ (or DCQ) facilitating the reaction to proceed from the oxyferrous form under high substrate concentrations.

TCP and TBP both occur as intermediates in industrial synthesis. TBP is more common, as it is used in the synthesis of brominated flame retardants and can then be released into the environment as a degradation product of these substances. The environmental and public health impact of these compounds has not been extensively explored as yet. Enough is known to cause growing concern around their levels in aquatic environments and their impact on male foetal development in humans and other mammals [188, 189]. The concern around these pollutants makes them a target for bioremediation with a robust peroxidase.

1.8 Classical peroxidase mechanism and kinetics

Much of the early work on understanding the mechanisms of peroxidase kinetics was performed on HRP. A green intermediate (compound I) of HRP was isolated as early as 1942 [190]. A three-step reaction mechanism with two single electron oxidations of substrate was then proposed in 1970 [191]. Due to the structural and mechanistic similarities between peroxidases, many results taken from one heme peroxidase can be applied to another. The equations below show the sequential reaction of HRP and two equivalents of ferrocyanide ($Fe(CN)_6^{4-}$) as an electron donating substrate but could be generalized to any peroxidase and electron donating substrate.

$$\mathrm{HRP} + \mathrm{H_2O_2} \xrightarrow{k_1} \text{Compound I} + \mathrm{H_2O} \quad (1.3)$$

$$\text{Compound I} + \mathrm{H^+} + \mathrm{Fe(CN)_6^{4-}} \xrightarrow{k_2} \text{Compound II} + \mathrm{Fe(CN)_6^{3-}} \quad (1.4)$$

$$\text{Compound II} + \mathrm{H^+} + \mathrm{Fe(CN)_6^{4-}} \xrightarrow{k_3} \mathrm{HRP} + \mathrm{Fe(CN)_6^{3-}} + \mathrm{H_2O} \quad (1.5)$$

This scheme is made up of experimentally observed steps. Each of the single electron reductions (1.4 and 1.5) must be accompanied by the addition of a proton if the peroxidase compound II is a protonated species [192]. However, there is conflicting evidence around this point. This will be discussed later in the chapter.

1.8.1 Formation of Compound I

The formation of compound I (1.3) experimentally appears to be a second order process with a typical rate constant of near 10^7 $M^{-1}s^{-1}$ [193]. Derived in HRP by monitoring pseudo-first order decay at 405 nm with saturating enzyme concentration and varying H_2O_2 concentration. In reality, compound I must be formed through an enzyme-substrate complex which is known as compound 0. Equation 1.7 shows the relationship between k_1 and the microscopic rate constants involved in the formation of compound I (1.6) [152].

$$\mathrm{HRP} + \mathrm{H_2O_2} \underset{k_b}{\overset{k_a}{\rightleftharpoons}} \underset{\text{Compound 0}}{\mathrm{HRP{\cdot}H_2O_2}} \xrightarrow{k_c} \text{Compound I} \quad (1.6)$$

$$k_1 = \frac{k_a k_c}{k_b + k_c} \quad (1.7)$$

The first widely discussed mechanism for the formation of Compound I in the literature is the Poulos-Krant "dry" mechanism [194]. This was predominantly based on observations from the initial X-ray structures and mutation studies that showed a 10^5 reduction in Compound I formation by removing the dHIS in CcP [145, 194, 195]. In the dry mechanism the dHIS serves as an acid-base catalyst by facilitating the transfer of a proton directly back and forth from the peroxide. A drawback to this mechanism is that the distance seems too great for direct interactions between the dHIS and the hydrogen of peroxide closest to the heme (H_1). It was also thought to be unlikely that the peroxide is deprotonated before the O-Fe association as the pKa of peroxide is near 11.6. Initially, computational calculations supported the dry mechanism showing that the H_1 of peroxide being deprotonated by the

dHIS to form ferric-hydroperoxide (compound 0) is a lower energy pathway when compared to moving the proximal oxygen to the distal oxygen (1,2- proton shift) [196]. Even the lower energy dry mechanism from this 2006 Shaik *et al*. study was far higher than the experimentally predicted for HRP (1.6 ± 0.7 kcal/mol) [197].

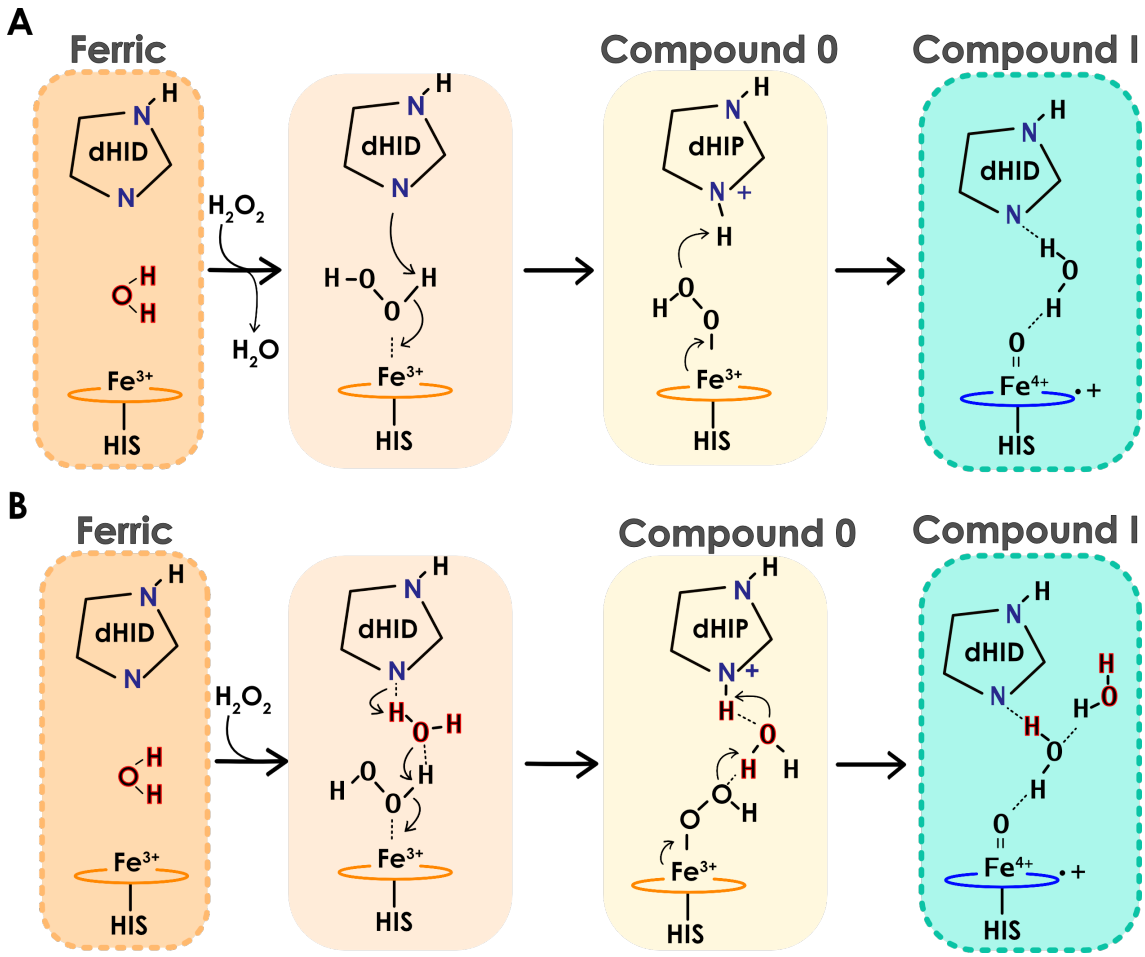


Figure 1.8.1 Compound I formation mechanisms. ***(A)*** *shows the "Dry" mechanism of compound I formation* [196]. ***(B)*** *shows the "Wet" or water mediated mechanism* [198, 199]

Jones and Dunford in 2005 postulated that the dHIS pKa must be increased to make it a viable proton acceptor to deprotonate H_2O_2. This was inspired by the observation that the rate of compound I formation is pH independent from pH 2.5 to 11. They postulate that the pK_a of the dHIS is shifted from around 2.5 in the resting state to near 11 during compound 0 formation. One hypothesis was that an additional water molecule relays a proton and mediates an electrostatic influence on dHIS in HRP [200]. Using this as inspiration, further calculations with a mediating water during compound I formation gave more favourable

energy barriers [198]. The wet mechanism for compound I formation in HRP was then further substantiated by Vidossich *et al*. with a combined MD and QM/MM analysis in 2010 [199]. Poulos recently updated the original Poulos-Krant mechanism to the new water-mediated mechanism explored by these computational studies [119].

In 2014 the neutron structures of CcP compound I by Casadei *et al*. challenged the generally assumed peroxidase mechanism. They reported the dHIS to be doubly protonated ($dHIS^+$) in compound I of CcP [5]. As a result, any mechanism to form compound I must require an additional proton from the solvent or another source. Both protons from peroxide may not be used to form water. An additional proton must be released between compound II and resting ferric states. Casadei *et al*. suggested mechanisms that seem to favour non-water mediated proton transfer between the dHIS and peroxiferric oxygen.

The Poulos *et al.* responded to the results of the neutron structure reiterating their support of a water-mediated mechanism [201]. They suggest that the dHIS in CcP compound I has a pK_a around pH 8. This could then be explained by protonation from the solvent under the crystallisation conditions of the neutron structure (pH 6). They performed two calculations: one used the H++ webserver (biophysics.cs.vt. edu/index.php); another compared the $\Delta G_{protein}$, changing protonation state of a residue in the protein environment with the ΔG_{water}, the titration in solution [201]. The ΔG_{water} they used standard results supplied with AMBER12 calculated with a Generalized Born implicit solvent model. $\Delta G_{protein}$ was calculated using Monte Carlo sampling of a 2 ns MD simulation with the backbone of the protein fixed in the same implicit solvent implementation. The pK_a predicted for the resting state was 5.1 and for compound I 7.8. These are broadly in line with resonance Raman studies of HRP isoforms that suggest that the dHIS has a pK_a of around pH 8.6 and interacts with the heme environment through the solvent [202].

The role of the dARG in CcP has been shown to be mechanistically important but not essential. A dARG to Leucine CcP mutant had a 10^2 factor reduction in the rate of compound I formation [203]. It was initially proposed just to form H-bonds to stabilise compound 0 in

the formation of compound I, but with the first crystal structure in 2003 of the CcP compound I, its role was expanded to stabilise compound I itself. These structures showed that the dARG was in two conformations "out" in the ferric enzyme and "in" hydrogen binding to the compound I ferryl oxygen [204]. After the neutron structure showing compound I with a $dHIS^{+}$ it has been suggested that dARG may mediate proton transfer from the solvent during the formation compound I and reduction [205].

1.8.2 Compound I radicals

The first compound I intermediate to be observed was in HRP [190]. It was first seen as a green intermediate in the 1940's and later assigned as a π-cation porphyrin radical [206]. W191 (pTRP) was conclusively identified as the radical site in CcP compound I by Electron nuclear double resonance (ENDOR) combined with isotopic labelling of either methionine or tryptophan including a W51 (dTRP) knockout [156]. A year earlier it had been shown that removing pTRP from CcP showed a 3000-fold decrease in rate of turnover [207]. The neutron cryo-crystallography structure of CcP compound I reveals a short Fe=O bond that is not protonated [5]. The dHIS doubly protonated (HIP) and is not directly hydrogen bonding with the ferryl oxygen. The dARG is in the "in" position, hydrogen binding to the ferryl oxygen [5, 204]. The crystal structure of APX revealed a tryptophan residue in the same proximal position as found in CcP. It was assumed that AXP would also produce a tryptophan radical. However, both freeze-quench EPR and stopped flow spectroscopy identified APX compound I as a porphyrin radical [208, 209]. The porphyrin radical is found in most other general peroxidases like HRP which has phenylalanine residues in place of the tryptophan residues found in APX and CcP.

The protein environment of AXP and CcP are subtly able to electrostatically stabilise different radical cations. Mutation studies of CcP, removing pTRP, show the preferential binding of positive small molecules into the newly formed pocket [210]. This, along with results from computational models, support the view that protein environment of this pocket stabilises the $pTRP^{+\bullet}$ positive charge in CcP [211]. Crystal structures of the plant peroxidases

APX, HRP and peanut peroxidase, have a cation binding site (K^+ or Ca^+) close to the proximal tryptophan (8Å in AXP) [148, 149, 212]. This K^+ binding site in AXP has been shown to be structurally important and seems to play a role in destabilising the $pTRP^{+\bullet}$ [213]. A K^+ binding site can be engineered into CcP resulting in poor activity and a loss of the characteristic $pTRP^{+\bullet}$ radical EPR signal [214]. CcP methionine 230, 231 and 172 have all been also implicated in stabilising the $pTRP^{+\bullet}$ [215]. It is possible that there are many alternate radical sites in CcP owing to the unusually high content of aromatic residues for a 30 kDa protein: 7 tryptophans and 14 tyrosines. It is possible (and highly likely) that the radical site can migrate amongst these residues. CcP forms tyrosine crosslinks (on residues 32 and 48 on surface) in the absence of electron donor [216]. This suggests that the oxidation of these sites could also be preferable to oxidation of the porphyrin.

1.8.3 Compound II

The structure of AXP compound II has been solved by neutron diffraction and shows the dHIS to be doubly protonated ($dHIS^+$) and the heme intermediate to be a Fe^{4+}-OH [6]. A protonated compound II oxygen is consistent with the other heme enzyme CYP-158, a bacterial P450, which has been observed to have an unprotonated compound I (Fe^{4+}=O;$Porphyrin^{\bullet+}$) but a protonated compound II (Fe^{4+}-OH) [217, 218]. This is the same trend observed in aromatic peroxygenases and chloroperoxidases [219–222]. A typical heme Fe^{4+}=O bond length is 1.65 Å, and in Fe^{4+}-OH, the Fe-O bond would typically fall between 1.76-1.81 Å. For Fe^{2+}-OH and Fe^{3+}-OH, typical Fe-O bond lengths would be 1.81-1.85 Å and 1.85-1.92 Å respectively [223]. A number of techniques have been used to estimate the pK_a of compound II in HRP, which is predicted to be less than 4. If this is the case, it would normally be deprotonated at HRP optimum pH 4.6. The conflicting evidence from different peroxidases means that it is still not known if CcP is an outlier amongst peroxidases.

1.8.4 Reduction of compound I and II

The reduction of compound I occurs rapidly in the presence of an electron donating substrate. This step is generally considered to be rapid, with the reduction of compound II

considered the rate-limiting step in most conditions [224]. In peroxidases where substrates typically associate at the heme edge, this is perhaps a result of the different electron transfer distances between Compound I and II. In compound I reduction the electron travels to delocalised the porphyrin radical but in the reduction compound II the electron must travel further to reduce the heme iron[152]. Compound II is always reduced with proton-coupled electron transfer. The donating substrate is often the source of this proton but this is not always the case. For HRP, the compound II reduction is dependent on a distal amino acid sidechain with a pK_a near 8.6, and above this pH it is inactive. This group is generally assumed to be the dHIS. The reduction of compound II is known to be the rate limiting step.

1.8.5 Comparison of Peroxidases and P450s

P450s and peroxidases access the same reactive intermediates (Compounds 0, I and II) but conduct quite different chemistries. The differences in the heme environment between these two families of heme oxidoreductases can help to rationalise the differences in reactivity. The topology of the heme cavity in P450s is larger and more hydrophobic and allows the binding of substrate close to the heme iron [119]. The opposite is true for peroxidases as their distal heme cavities are more restricted and substrates are force to bind at the heme edge.

The cysteine ligand of P450s is more electron donating than the histidine in peroxidases. The thiolate-iron ligation must generate an oxidant strong enough to break the 100 $kcal.mol^{-1}$ C–H bond of a typical hydrocarbon [144]. The difference between peroxidase and oxygenase activity is thought to be due to the pK_a of the Fe(IV)-OH hydoxyferryl species (compound II). P450s generally possess a long Fe-O bond in compound II - around 1.82 Å - where in peroxidases it is typically 1.7 Å. The more electron donating thiolate ligation in P450s results in a basic pKa ferryl oxygen (above 8.2) that supports hydrogen bond abstraction [144].

1.9 Computational description of proteins

Computational models allow the relationship between structure and dynamics to be explored at the atomic level. Quantum mechanical (QM) methods involve solving approximately the electronic Schrödinger equation to determine the wave function and energy

of a system. In practice this in-depth description of electronic structure can only be calculated for systems with a small number of atoms around 200 [225]. For proteins, made up of many thousands of atoms, molecular mechanics (MM) methods are instead employed [226]. The results chapters of this project include molecular dynamics (MD) calculations which solve a MM description of the system through time. Chapter 4 discusses work that combines QM and MM calculations. An introduction to the QM method used is discussed in Section 5.1.

1.9.1 Molecular Dynamics

Molecular mechanics methods assume that the energy of the system can be calculated solely by the nuclear positions, disregarding the computationally expensive electronic components. Atoms are modelled as small particles and bonds as connecting springs [227]. The system is described by classical mechanics (MM, Newtonian equations of motion) using potential-energy functions for inter- and intra-molecular forces. This description of the system is known as the force field. It contains generalised descriptions of bonded interactions (bond: stretching, bending, rotation) and non-bonded interactions (electrostatic and Van der Waals interactions) for all the different atom types in the system (Figure 1.9.1) [225].

Results using a force field, rigorously parameterised by experimental and QM data, can in some cases produce structures as accurate as QM calculations but take a fraction of the computational resources. Molecular dynamic (MD) simulations build up a dynamic picture of a molecule by applying the MM description of the system over time. It offers a compromise between accuracy, validity and practicality [227]. Two widely tested protein force fields were used in this project (CHARMM22 and AMBER14ffSB) across two different software packages: AmberTools15 and Gromacs 4.6.4 [228–231].

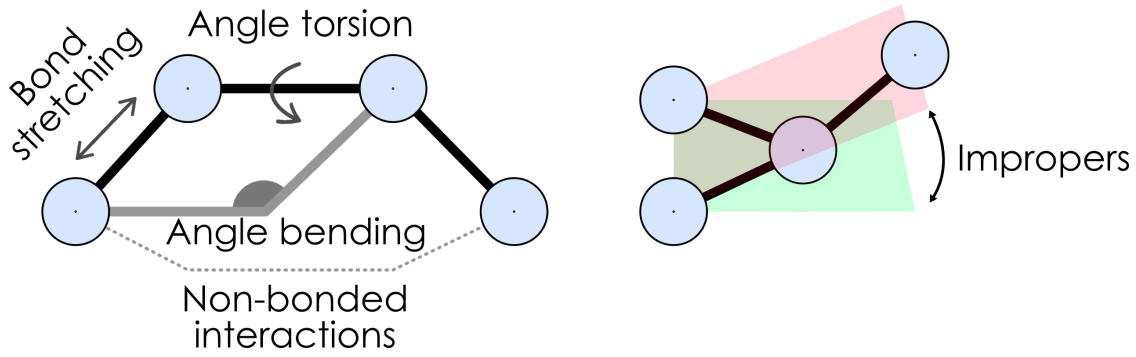


Figure 1.9.1 MM energy terms.
Illustration of different energetic terms parameterized in a MM force field

Aside from the inherent approximations and assumptions of the force fields method, the largest source of uncertainty lies in the validity of the starting structure. Generally, the length of MD simulation limits the conformational search of the starting structure, only allowing the simulation to converge in a local energy minimum [227]. This means that the less accurate the starting structure the less representative the MD simulation will be of reality. The current computational model used for the maquettes in this project is not directly based on an experimental structure, as there is not one available for C45. However, a related protein was crystallised in 2003, which shares 45% sequence identity [65]. Experimental observations with circular dichroism spectroscopy show that C45 is almost completely helical. The computational structure of C45 is based on the assumption that it folds to a similar gross structure of the crystallised maquette.

The different conformations sampled by the protein during the MD simulation can then be used as a starting structure for QM calculations and ligand docking calculations. When combined with docking experiments, MD simulations can give information on the structural stability of the enzyme and interactions with small molecules.

1.10 Aims

This project focuses on *c*-type heme binding maquettes that have been designed to mimic the functionality of peroxidases *en route* to the more commercially interesting oxygenase or peroxygenase chemistry. Chapter 2 aims to explore the final design iteration that produced a mono-histidine heme ligation in a *c*-type heme maquette, named C45. Biophysical, kinetic

and computational characterisation of the peroxidase maquette will be employed with the aim of furthering our understanding of the minimum structural requirements for peroxidase activity.

The third chapter aims to study the reactive intermediates generated during the peroxidase cycle of C45. In particular, the presence or absence of a tryptophan protein radical and its effect on the functionality of the peroxidase. The fourth chapter aims to complement the bottom up approach to studying peroxidase activity by using high level computational modelling to study the reactive intermediates of natural peroxidases (CcP and AXP). The dual purpose of these calculations is to both shed light on a contentious issue in the natural peroxidase mechanism and to investigate computational techniques that could be used in the future to inform the design of *de novo* heme protein maquettes.

Chapter 2

Characterisation of the *de novo* peroxidase

2.1 Introduction

The maquette design process has been successful in producing functionalised *de novo* proteins. Section 1.3 details the progression of the maquette protein discussed in this chapter from a short four-helix bundle (4HB) produced on a peptide synthesiser, to longer proteins looped together and capable of binding cofactors [39, 63]. Heme B binding maquettes could modulate the redox properties of their cofactors and could be designed to mimic the functionality of haemoglobins in reversibly binding oxygen [61, 62]. More recently the design of *c*-type heme maquettes has allowed for the cofactor to be covalently incorporated *in vivo* allowing for greater control of the orientation and positioning of the heme within the bundle [69].

The aim of creating the heme binding maquettes is to reproduce and potentially exceed the sophisticated enzymatic activity found in natural heme oxidoreductases. The first target is peroxidase activity, *en route* to oxygenase chemistry as a long-term goal. Natural peroxidases such as HRP, CcP, AXP and LiP all have heme in a pentacoordinate ligation scheme [145, 147–149]. The distal site is then free to bind water in the resting state and for H_2O_2 to start the reactive cycle [119]. Removing one of the heme-ligating histidine residues in the heme B binding maquettes severely diminishes heme binding. For *c*-type heme only a single histidine is required for binding in the CXXCH biding motif. The cysteine residues of the recognition

sequence are covalently bound to the heme by the endogenous Cytochrome *c* biogenesis system I (Ccm) in *E. coli* [71].

The aim of this chapter was to produce a mono-histidine ligated maquette and then characterise any resulting enzymatic activity. Covalent heme binding in *c*-type maquettes has unlocked the possibility of removing the distal histidine of the heme without a loss in heme binding. The design rational was to mimic natural peroxidases where the distal site of the heme is free to interact with solvent. The *c*-type heme maquette produced, C45, is the first peroxidase maquette, indeed the first maquette to exhibit enzymatic activity on the same order of magnitude as natural enzymes. This chapter is a detailed characterisation of the structure, kinetic activity, substrate binding and substrate promiscuity. The work is essential for understanding what the minimal requirements for producing peroxidase activity in a *de novo* scaffold are.

2.2 Materials and Methods

2.2.1 Mutagenesis and Cloning

The mutagenesis of the maquette variants was performed using a Site-directed, Ligase-Independent Mutagenesis (SLIM) protocol [232]. Both the SLIM protocol and subsequent PCR experiments were performed using Q5 High-Fidelity DNA Polymerase. The proteins were expressed from a pMal-p4x vector with the maltose binding proteins removed. The sequences encoding the Sec signal sequence, a Histidine affinity tag and the tobacco etch virus (TEV) cleavage site were all placed directly upstream of the maquette gene. The primers used in the mutagenesis were designed using the programme CloneMUTS2 and synthesized by Eurofins (MWG, UK) [233].

After mutagenesis, a QIAGEN PCR cleanup kit was used to prepare the plasmid for transformation. 50 μl of *E. coli* Stellar competent cells (New England Biolabs, UK) were transformed by 42°C heat shock for 45 seconds after a 30 minute incubation with 1 μl of plasmid on ice. After heat shock, cells were then left on ice for 2 minutes before adding 450 μl of super optimal broth (SOC, [2 % tryptone, 0.5 % yeast extract, 2.5 mM KCl, 10 mM NaCl and MgSO4, 20 mM glucose]) as a growth medium. The transformed cells were incubated at 37 °C for 1 hour before pelleting and spreading onto 1 % agar lysogeny broth (LB) plates containing 50 μg/ml carbenicillin.

After overnight incubation, selected colonies were used to inoculate 5 ml of LB (50 μg/ml carbenicillin) and incubated overnight with shaking. Samples were centrifuged at 4000 xg for 5 minutes to harvest the cells. Plasmid DNA was extracted using a standard miniprep kit (Sigma, UK) and sequenced by Eurofins. Two steps on the standard miniprep protocol were modified to increase plasmid yield. The plasmid DNA was bound to the column with a slow spin (800 rpm for 2 min, followed by a pulse to 13,000 rpm) and the plasmid was eluted with water at 70 °C.

Correctly mutated plasmid DNA was co-transformed into *E. coli* T7 express cells (New England Biolabs, UK) with a pEC86 constitutive expression vector housing the Cytochrome *c* biogenesis system I genes and chloramphenicol resistance [234]. The co-transformation only deviates from the above method in that cells incubated with 1 μl of both plasmids for 30 minutes on ice before a 10 second heat shock. The mutagenesis was determined to have been successful by sequencing the relevant portion of the plasmid (Eurofins, MWG).

2.2.2 Protein expression and Purification

Glycerol stocks of successfully co-transformed *E. coli* were used to inoculate 50 ml of LB as a starter culture. All LB used in the expression contained carbenicillin (50 μg/ml) and chloramphenicol (34 μg/ml). After incubation with shaking overnight at 37 ˚C, starter cultures were added to 1 L of aerobic LB. The 1 L culture was incubated in the same conditions. When the optical density of the cultures at 600 nm reached 0.6-0.7, maquette expression was induced with the addition of 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG, Apollo Scientific). 4 hours of incubation (37 ˚C with shaking) completed the expression phase of production. The cells were pelleted by 4,000 xg centrifugation and resuspended in lysis buffer (300 mM NaCl, 50 mM NaPhos, 20 mM imidazole, pH 8.0).

Phenylmethanesulfonylfluoride (PMSF), a serine protease inhibitor, was added to 1 mM and the lysate kept below 4 ˚C when handling to avoid proteolysis. Three cycles of Sonication (using Soniprep 150, MSE UK) at 100% amplitude with 30 second pulses separated by 30 seconds of mixing lysed the cells pellets. The aqueous and lipid phases of the lysed sample were subsequently separated by centrifugation at 40,000 xg for 30 minutes.

A 5 ml Nickel affinity chromatography column (HisTrap HP, GE Healthcare, UK), equilibrated in lysis buffer was loaded with the aqueous lysate (filtered through 0.2 μm pore filter). A red band showing the his-tagged, heme-containing maquette clearly binds to the column whilst the vast majority of cellular proteins flow through. The histagged maquette was eluted with an increasing gradient of elution buffer (300 mM NaCl, 50 mM NaPhos, 250 mM imidazole, pH 8.0). The fractions containing maquette were collected and dialysed into TEV cleavage buffer (20 mM Tris, 0.5 mM EDTA, pH 8.0) in 8,000 MW tubing overnight.

The N-terminal His-tag was cleaved during a 4 hour incubation (room temperature in anoxic conditions) with TEV protease (100 μg/ml) and tris(2-carboxyethyl)phosphine (TCEP, 1 mM). The sample was then loaded onto a nickel affinity column in the same conditions as the first nickel column, and the flowthrough is harvested. The sample was then loaded onto a HiLoad Superdex 75 pg size exclusion column after being concentrated to approximately 4 ml using a spin concentrator (Vivaspin 20, 5,000 MWCO). The column was then equilibrated with redox buffer (100 mM KCl, 20 mM CHES at pH 8.6). The maquette eluted around 43 minutes after injection with a 1 ml/min flow rate. Samples were taken

throughout the purification to perform sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS -PAGE).

The pure samples were then concentrated to a useful concentration, above 100 μM. A pyridine hemochrome assay was used to accurately determine the concentration of heme-containing protein. The assay was performed in a low volume cuvette. 90 μl of diluted heme-protein sample (typically and 75 times dilution) was added to the cuvette and spectra recorded. Then 90 μl of a solution containing 40 % pyridine and 0.4 M NaOH in the same buffer as the sample (redox buffer pH 8.6) was added. The solutions were mixed thoroughly for 15 seconds (30 aspirations) and the oxidised spectra recorded directly after. Immediately after this the sample was reduced by the addition of sodium dithionite (a small but saturating quantity of powder). The sample is mixed once more for 15 seconds on the addition of sodium dithionite and the reduced spectra taken immediately after. The concentration of heme was calculated by the difference in oxidised and reduced spectrum ($\boldsymbol{\varepsilon}\,\mathbf{A}_{550\,red}$ - $\mathbf{A}_{550\,Ox}$ = 21840 $M^{-1}.cm^{-1}$) [235].

2.2.3 UV-Visible Spectroscopy

A Cary 60 spectrophotometer (Agilent Technologies) was used to record all UV-visible spectra. Reduced (ferrous) spectra were measured after the addition of sodium dithionite in excess.

2.2.4 Circular Dichronism

Circular Dichronism (CD) was used to examine the secondary structure of the protein maquettes. The temperature dependence of folding was assessed using a J-1500 spectrometer (Jasco, UK). Experiments used 300 μl of 10 μM protein in 1 mm path length quartz cuvette sealed with a stopper. For the melt experiments the samples were heated from 5 to 95 ˚C at a rate of 40˚ C/hour. At each degree of temperature change the ellipticity at 222 nm was recorded to assess the α-helical content of the protein. At 5, 25, 35, 55, 75 and a 95 ˚C full spectrum scans from 250 to 190 nm were taken. Cooling experiments were performed on the heated sample with the same conditions but in reverse. The ellipticity output (mdeg) was converted to mean residue ellipticity ($deg.cm^2dmol.res^{-1}$) using Equation 2.1.

$$\theta = \frac{E}{10(P.c)l} \tag{2.1}$$

In Equation 2.1: "θ" is the mean residue ellipticity, "E" is the ellipticity in mdeg, "P" is the number of peptide bonds (number of residues – 1), "c" is the molar concentration of the protein (M) and "l" is the path length in cm. Melts with 2, 4 and 6 M Urea were also performed in the same way.

2.2.5 NMR

The maquette protein expressed from *E. coli* grown in ^{15}N-enriched media was examined by 1D ^{1}H and 2D ^{1}H -^{15}N-HSQC experiments on the BrisSynBio Cryo 700MHz (Bruker, UK) instrument at The University of Bristol. The same growth and purification protocol explained in Section 2.2.2 was used except for the addition of 1 g/L ^{15}N-ammonium chloride before induction for 4 hours. A preparation of ^{15}N-enriched C45 was produced with the heme iron removed. For this 1.5 ml of Hydrogen fluoride-pyridine (70% HF) was added to lyophilized C45 in a 50 ml Falcon tube (Polypropylene), and stirred for 10 minutes. The resulting metal-free protein was light sensitive and so was kept in the dark. The reaction was quenched by addition of 10 ml of 50 mM ammonium acetate (pH 5.0). Then Nitrogen was flowed over the sample for 2 hours before buffer exchange and 4 h of dialysis against 10 mM sodium acetate (pH 5.0) [236]. A 60 μl sample of 200 μM was then examined by 1D proton and 2D ^{15}N-HSQC experiments.

2.2.6 Molecular Dynamics

The majority of the computational work was run on the University of Bristol's supercomputer, BlueCrystal phase 3. Both AMBER and GROMACS MD simulations used the same pdb coordinate file as a starting structure. Each was simulated in a cuboidal periodic box (> 18 Å from protein) with TIP3P explicit solvent and with 150 mM sodium and chloride ions added. The histidine residue in the CXXCH motif of C45 was singly protonated on the Nδ (HID). All other ionizable residues were in their physiologically relevant protonation states. The trajectories from both simulations were visualised and analysed in VMD [237]. The structural identity (Q_H value) was determined using Multiseq [238].

2.2.6.1 GROMACS 4.6.4

MD of C45 in GROMACS was built and minimised on a local Linux machine. The pdb file was converted into a structure file (.gro) and a topology file (.top) in the GROMACS format. Then the periodic boundary conditions were added, the box solvated with TIP3P water

and ions were also added. The CHARMM27 force field was used to parameterise the system, and the covalently bonded heme and histidine-iron coordination bonds were specially parameterised in a separate file [239]. For C45 the CHARMM27 force field was modified to include a set of *c*-type heme parameters derived from the standard CHARMM heme B parameters. The initial energy minimisation was run with no restraints and a steepest descent algorithm that converges when the maximum force is smaller than 200 kJ mol^{-1} nm^{1}. Second, a pressure equilibration was performed for 200 ps without generating initial velocities. For both the second minimisation and the production MD an NPT ensemble was used with verlosity-rescale temperature coupling and isotropic Berendsen pressure coupling at 300 K and 1 atm . The production MD was run on Bluecrystal at 2 fs steps producing in total a 100 ns trajectory written out into 100 frames (saving every 100 ps). The results were processed and water removed with trjconv and then visualised and analysed in VMD [237].

2.2.6.2 *AMBER*

It was important to convert the MD setup into AMBER as the output files from AMBER simulations are readily compatible with Chemshell and ORCA, programmes used to perform QM calculations [240, 241]. The same initial coordinate file as the GROMACS simulations and the same CHARMM22 force field (the protein portion of the CHARMM27 force field) were used [228, 239]. Empirically-derived topologies and parameters for *c*-type heme in the AMBER format were helpfully provided by the Luthey-Schulten group [242]. The MD was setup on a local machine using VMD to build a TIP3P water box with ions. VMD was also used to patch together the protein and *c*-type heme parameter, topology and structure files into a coordinate file (pdb) and protein structure file (psf). Chamber within parmed.py (part of AMBERtools16) was used to produce AMBER compatible parameter-topology and coordinate files (.prmtop and .inpcrd). Finally, before running the energy minimisations, atom names of the TIP3P water had to be edited to make them compatible with AMBER naming conventions. A number of preliminary energy minimizations were performed; a minimisation of solvent and hydrogens with protein heavy atoms restrained, a heating of solvent with restraints for the protein, a minimisation of the entire system, a quick heating of the entire system and a final 200 ps NPT ensemble MD equilibration of the system (details in Appendix B). The first minimisation was performed on a local machine with AMBER16 installed whilst the remaining minimisations and the production MD were calculated on a

graphics processing unit (GPU) node of Bluecrystal phase 3. The production MD was run at 2 fs time steps for a total length of 100 ns. A weak-coupling algorithm to maintain a constant temperature at 300° K and isotropic Berendsen pressure coupling at 1 atm were used [243]. Structures were saved every 100 ps. The system was initially run with a non-bonded cutoff of 12 Å (the AMBER default is usually 8 Å) so that the Amber and GROMACS results could be more directly compared. Cpptraj in AMBER tools and VMD were used to analyse and visualise the results.

2.2.7 Preliminary computational folding with Rosseta

Rosetta 3.8 (rosettacommons.org) was downloaded and compiled on a local Linux desktop running Ubuntu 16.04 LTS (Ubuntu.com). To run the folding calculations, the AbinitioRelax.default.linuxgccrelease program was used [30]. All Rosetta *ab initio* calculations require an amino acid sequence in fasta format and fragment files that are generated using the ROBBETTA webserver (robetta.bakerlab.org). These are short structural motifs mined from the Protein Data Bank (PDB) of either 3 or 9 amino acids long. Two different conditions were run: one with and without the heme. Calculations without heme were setup with the default setup parameters and can be found in Appendix B1.

Simulations with heme required additional Rosetta parameter files: full-atom heme residue parameters, centroid heme residue parameters, full-atom constraint file, centroid constrain file and residue pair jump constraint file were all employed. The constraint files contain additional harmonic restraints to force a *c*-type heme conformation of the CXXCH motif. The residue pair jump constraint file contains the instructions to link the non-standard residue to the structure as a rigid body. All these files were designed from scratch taking inspiration from a sparsely commented setup included with Rosetta, *ab initio* fold of a zinc chelating protein (found in main/test/integration/tests/metalloprotein_abrelax/). Whilst this setup was valuable in outlining the method, a large amount of trial and error was still required to produce a useful output, let alone something resembling a folded protein. The full atom and centroid heme residue parameters were based on a *c*-type heme mol2 file adapted from AMBER ferrous heme B parameters [244]. A python script bundled with the Rosetta installation (molfile_to_params.py) was used to generate the first template of these files which had to be manually edited to function as a *c*-type heme.

Whilst the development of the setup was performed on a local linux computer, the bulk of the calculations were run on the University of Bristol's BlueGem Synbio cluster.

2.2.8 Kinetic assays for ping-pong fitting

Ping-pong kinetics with ABTS and H_2O_2 were carried out using KinetAsyst SF-61DX2 Stopped-Flow Spectrophotometer (Hi-Tech Limited, UK). Six concentrations of ABTS (3, 2, 1, 0.5, 0.2, 0.1 mM) were all tested against six concentrations of H_2O_2 (5, 10, 25, 50, 100, 150 mM) monitoring at 525 nm (ABTS ε_{525nm}= 4250M^{-1} cm^{-1}). The dynafit software package was used to fit the data to a Ping-Pong mechanism. The data was recorded in triplicate with Dr Katie Grayson and fitting was performed with the help of Dr Gus Cameron, as cited in [70].

2.2.9 pH-dependent kinetics and spectroscopy

The buffers used were: 20 mM Potassium Phosphate (pH 6 to 8), 20 mM CHES (pH 8.5 to 10), and 20 mM CAPS (pH 11). The same protocol as described for kinetic assays with ABTS as described in Section 2.2.12 was used except all reagents were made up in the relevant buffer for pH. The pH dependent Optically transparent thin-layer electrochemistry (OTTLE) was performed by the same method as seen Section 3.2.7 except the relevant buffer for the pH rage as (described above) was used.

2.2.10 Temperature-dependent kinetics

Temperature dependent kinetics were performed in 100 μl reaction volumes with 0.25 μM C45, 100 μM H_2O_2 and 0-100 μM ABTS concentrations. Spectra were recorded with a Cary-60 UV-visible spectrophotometer (Agilent Technologies) using a TC 1 peltier (Quantum Northwest) and BATH 10 circulator (Quantum Northwest) to control the temperature in a t2 Sport/Cary 60 cuvette holder (Quantum Northwest). 90 μl of ABTS and C45 in redox buffer at pH 8.6 was equilibrated for 5 minutes at either 20, 30, 40, 50, 60 or 70 ˚C. A new stock of 1 mM H_2O_2 was prepared for each temperature and equilibrated. Reactions were recorded in triplicate and initiated by mixing 10 μl H_2O_2 into the solution of C45 and ABTS. These experiments were performed by Daniel Watkins as cited in [70].

2.2.11 Computational binding screen with BUDE

BUDE (The Bristol University Docking Engine) surface scanning mode was used for 19 ligands on 10 different snapshots from a 100 ns GROMACS MD simulation of C45 [245]. The snapshots were taken every 10 ns of the simulation. The structures of the ligands were

taken from the Zinc database online and parameterised with the programme acpype.py using Antechamber and the General AMBER Force Field (GAFF) [246–248]. BUDE only requires mol2 coordinates and autotypes as input for ligands. A single conformer of each ligand was used except for ABTS where 10 different conformers were used, generated by a previous MD simulation. Charges of ligand functional groups were set to their value at pH 8.6 (see Figure 2.3.17). The BUDE setup, BUDE running, processing for MD, MD setup and MD post-processing was scripted where appropriate. Each combination of ligand and snapshot of C45 simulation (receptor) had 1,200 randomly assigned surface positions tested. At each of these positions 3,000 conformational iterations of the starting ligand pose were tested with the top 30 energies selected to be parents. These parent poses were then randomly iterated a further 3,000 times and a new generation of top 30 energy parents selected. This process was repeated for ten generations at each starting pose. In total 1.08×10^{10} conformations are sampled per ligand with C45, with each generation following a path to stronger binding. The top 50 binding modes per ligand were then individually processed and run as a 5 ns MD simulation using the AMBER setup, as previously described.

VMD was used to process the trajectories aligning all frames from across all ligands by the protein backbone and then calculating the centre of mass for each ligand with respect to the heme iron. For each frame, if the heavy atoms of the ligand were found to be over 5 Å from any part of the protein it was classified as dissociated. This information was then analysed using python, where a three-dimensional cluster analysis was performed on both the BUDE input structures and the MD results after 5 ns, separately. This was to give an approximate quantification of possible binding hotspots. Each of the start and end data sets analyzed the centre of mass of all ligands together. The cluster analysis used sklearn DBSCAN and comp [249]. All ligands marked as dissociated were binned in one cluster for the 5 ns analysis. VMD was also used to generate lists of close interactions (<3 Å) between the heavy atoms of the ligands and protein. The frequency of these interactions on the final structure of the 5 ns MD simulation were binned per protein residue and per protein atom and used to produce ligand binding heat map figures. An example of the BUDE setup, running and processing for MD script is found in the appendix (Appendix B)

2.2.12 Kinetic assays for various substrates

Kinetic assay were performed with the following substrates: ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonicc acid)), 2,4,6-Trichlorophenol (TCP), 2,4,6-Tribromophenol (TBP), Amplex red, Guaiacol, 3,3'-dimethoxybenzidine, Luminol, *o*-Phenylenediamine, *p*-Aminobenzoic acid, Ascorbate, Veratrole alcohol (VA), and 1,4-Dimethoxybenzene. All substrates were purchased from sigma at high purity (>95%). All experiments were performed a pH 8.6 in CHES buffer at 25 ˚C (referred to as Redox Buffer). In general, stock solutions of the substrates were diluted to stock concentrations of 400 μM and 10 μM in Redox buffer prior to use.

Exploratory experiments were performed on Cary 60 spectrophotometer (Agilent Technologies) scanning from 200 to 800 nm every 6 seconds and a 1 cm path length. These were performed at either 80 or 100 μM of electron donating substrate, 0.1 μM enzyme and 100 μM peroxide concentration. Reactions were initiated by mixing equal volumes of one solution containing enzyme and electron-donating substrate and another solution of peroxide.

Michaelis–Menten kinetic data were obtained using a KinetAsyst SF-61DX2 Double-mixing Stopped-Flow Spectrophotometer (Hi-Tech Limited, UK) with a 1 cm path length. Reactions were initiated by mixing equal volumes of one solution containing 0.1 μM C45 with various concentrations of electron-donating substrate, and a solution of 100 μM hydrogen peroxide. The initial linear portion of each trace was fit to a straight line and the gradient (ΔAbs.Sec^{-1}) determined using an interactive python script. The initial rates were converted to change in concentration over time using the relevant extinction coefficient for the substrate. Then these were fit to a Michaelis–Menten curve (equation 2.2) or substrate inhibition (equation 2.3) using GraphPad Prism.

$$y = \frac{k_{cat}\cdot[S]}{K_M+[S]} \tag{2.2}$$

$$y = \frac{ABS_{max}\cdot[S]}{[S]\left(K_M+\left(1+{}^{[S]}/_{k_i}\right)\right)} \tag{2.3}$$

3 mg of **ABTS** were dissolved as green powder into buffer, protected from light and kept at room temperature. This was made fresh for each use and diluted to make up stock required for kinetic experiments. In these conditions ABTS was stable over the course of 6 hours, however, some degradation was observed overnight. Oxidation of ABTS was

monitored at 405 nm (ε_{405nm} = 36,800 $M^{-1}.cm^{-1}$) [250]. ABTS concentrations used were 1, 5, 10, 20, 40, 80, 100, 150 μM.

1 mg of **TCP** crystals were dissolved in 100 μl ethanol and then diluted to 400 μM in redox buffer. The concentration was determined spectroscopically ($\varepsilon_{312\ nm}$ = 3,752 $M^{-1}.cm^{-1}$) [251]. In these conditions no degradation of TCP was evident, even after overnight incubation. Oxidation of TCP was monitored at 272 nm, monitoring the increase in dichloro-1,4-benzoquinon (DCQ) formation ($\varepsilon_{272\ nm}$ of DCQ - TCP = 13, 200 $M^{-1}.cm^{-1}$). TCP concentration used were 1, 5, 10, 20, 40, 80, 100, 150 μM [252].

40 mg of **TBP** (off-white solid pellet) was dissolved in 100 μl ethanol and then serially diluted to 400 μM in redox buffer. The extinction coefficients of the three UV-visible peaks of TBP were determined in redox buffer (316, 249 and 214 nm). This was achieved by recording the absorbance peaks at know concentrations of TBP in the range 1 – 150 μM, then using linear regression to determine the gradient. In these conditions TBP is stable over the course of 6 hours. Oxidation of TBP was observed at 249 nm (and 316 nm at substrate concentration above 70 μM) monitoring the decrease in substrate (ε_{249nm} 1,118 $M^{-1}.cm^{-1}$, $\varepsilon_{249\ nm}$ 4,994 $M^{-1}.cm^{-1}$). The concentrations used were 2.5, 5, 10, 20, 40, 80, 100, 150 μM.

Amplex red was purchased from ThermoFisher (UK). 10 mM stock solutions were prepared in the dark in DMSO and stored at -20 ˚C, as recommended by the supplier's protocol (Amplex Red Glutamic Acid/Glutamate Oxidase Assay protocol, Invitrogen). Amplex red was protected from light and kept on ice throughout the experiment. 400, 20 and 0.12 μM stocks of Amplex Red were made up in redox buffer. In these conditions little degradation of Amplex Red was seen after 6 hours. Oxidation of Amplex red was observed at 575 nm, monitoring the increase of the product resorufin (ε_{572nm} 73,000 $M^{-1}.cm^{-1}$)[253]. The concentrations used were 0.05, 0.5, 1.2, 2.5, 5, 10, 40, 75, 150 μM.

40 mg of **luminol** was dissolved in 200 μl 1 M KOH, then 800 μl deionized water was added to 1 ml. It was then diluted to 400 μM in redox buffer. The extinction coefficients of the three UV-visible peaks of luminol were determined in redox buffer (348, 302, 221 nm). This was achieved by recording the absorbance peaks at known concentrations of luminol in the range of 1 – 150 μM and using linear regression to determine the gradient. In these conditions luminol is stable over the course of 6 hours. Oxidation of luminol was observed at 302 nm (and 221 nm at substrate concentration above 70 μM) monitoring the decrease in

substrate absorbance (ε_{302nm} 6,830 $M^{-1}.cm^{-1}$, ε_{221nm} 27,760 $M^{-1}.cm^{-1}$). The concentrations used were 2.5, 5, 10, 20, 40, 80, 100, 150 μM.

1 μl **guaiacol** was diluted to 1 ml in deionized water and then diluted to 400 μM in redox buffer. The concentration was determined spectroscopically ($\varepsilon_{274\ nm}$ = 2550 $M^{-1}.cm^{-1}$) [254]. In these conditions guaiacol is stable over the course of the experiment. Oxidation of guaiacol was monitored at 470 nm monitoring the increase in tetraguaiacol formation ($\varepsilon_{470\ nm}$ = 3,800 $M^{-1}.cm^{-1}$) [254]. The guaiacol concentrations used were 1, 5, 10, 20, 40, 80, 100, 150 μM.

1 mg **o-dianisidine** was dissolved in 500 μl of deionized water. After the addition of 100 μl 1 mM KCl solution was made up to 1 ml. The concentration was determined spectroscopically in acid ($\varepsilon_{251\ nm}$ = 2,550 $M^{-1}.cm^{-1}$) and then serially diluted to 400 μM in redox buffer [255]. In these conditions o-dianisidine is stable over the course of the experiment. Oxidation of guaiacol was monitored at 440 nm monitoring the increase in oxidized o-Dianisidine formation ($\varepsilon_{440\ nm}$ = 11,300 $M^{-1}.cm^{-1}$) [255]. The o-dianisidine concentrations used were 1, 5, 10, 20, 40, 80, 100, 150 μM.

40 mg ***o*-phenylenediamine** was dissolved in ethanol and then diluted to 400 μM in redox buffer. In these conditions, *o*-phenylenediamine is stable over the course of the experiment. Oxidation of *o*-phenylenediamine was monitored at 450 nm monitoring the increase in phenazine-2,3-diamine formation ($\varepsilon_{450\ nm}$ = 16,300 $M^{-1}.cm^{-1}$) ([256] supplementary information). The o-phenylenediamine concentrations used were 1, 5, 10, 20, 40, 80, 100, 150 uM.

2.2.13 Live Cell kinetics

To investigate C45 *in vivo* activity preliminary experiments were performed with whole, live *E. coli* cells. As a control, a gene of C45 without a heme binding site was produced. Both the pSHT C45 and C45-NH (without a heme binding site) were co-transformed into *E. coli* T7 express cells (New England Biolabs, UK) with a pEC86 constitutive expression vector housing the Cytochrome *c* biogenesis system I genes and chloramphenicol resistance. The transformation was achieved by heatshock with a 10 second incubation at 42 ˚C. After incubation with shaking overnight at 37 ˚C 100 ml starter cultures were added to 1 L LB with carbenicillin (50 μg/ml) and chloramphenicol (34 μg/ml). Both cell cultures grew at similar rates and were induced with 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG,

Apollo Scientific) at an $OD_{600\ nm}$ between 0.6 and 0.7. After 3 hours of expression, activity assays were conducted in a 96-well format on a Synergy Neo2 Plate Reader (BioTek, UK). 30 μl of cells were diluted in 170 μl of electron-donating substrate in redox buffer (100 mM KCl, 20 mM CHES at pH 8.6). The reactions were initiated with a 100 μl injection of H_2O_2. The working concentrations of electron donating substrates was 1 mM, and H_2O_2 was at 2 mM in a total well volume of 300 μl. Volume-corrected spectra were obtained rapidly before injection, after 20 minutes and after 90 minutes. ABTS, Guaiacol, Luminol and *o*-phenylenediamine were used as electron-donating substrates and they were prepared in the same way as in the previous section.

As a further positive control, 500 μl of both cells expressing C45 and C45-NH were spun down in a microcentifuge and resuspended in 90 μl redox and 10 μl of BugBuster (Sigma, UK) to lyse the cells. 10 μl of lysate was added to 190 μl of redox buffer with ABTS. Reactions were initiated with the addition of 100 μl of H_2O_2 and monitored spectroscopically. The concentration of ABTS was 800 μM and H_2O_2 was 300 μM.

2.3 Results

2.3.1 Design and mutagenesis

The template used for producing C45 (the mono-his ligated *c*-type heme maquette) was a vector containing the parent protein C46 (the bis-histidine ligated *c*-type heme maquette). These vectors are modified pMal-p4x vectors where the maltose-binding protein gene has been deleted, leaving the MBP periplasmic translocation signal sequence to target the maquette for periplasmic export. Downstream of the cleavage site for the N-terminal signal sequence is a hex-his tag as well as the recognition sequence (ENLYFQ|G, where cleavage occurs between Q and G) for Tobacco Etch Virus protease (TEV). This facilitates removal of the his-tag during the purification of the maquette. Primers were designed to remove the distal histidine ligand from C46, replacing it with a phenylalanine (H56F mutation).

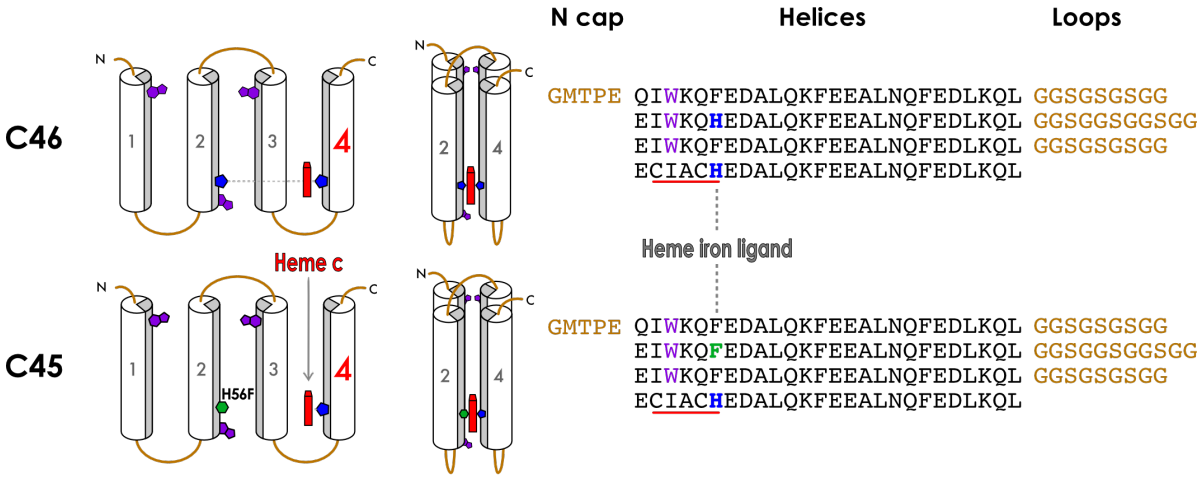


Figure 2.3.1 Schematic of C46 and C45

Figure showing the design of C46 and C45. The coloured residues in the protein sequence correspond to the coloured residues displayed in the schematics. The cylinders represent the helices of the maquettes with the grey shaded portion depicting hydrophobic residues. The red underlined sequences are the c-type heme binding motif CXXCH found on the 4th helix of both maquettes

2.3.2 Expression and purification

Protein expression proceeded as described in the methods Section 2.2.2. After four hours of expression the cell cultures were pelleted by centrifugation and resuspended in lysis buffer. Sonication was used to lyse the cells followed by clarification by centrifugation. Nickel affinity chromatography was used twice: first collecting the elution containing his-tagged C45, and second collecting the flow-through, after the cleavage of his-tag by TEV protease. Size exclusion chromatography (SEC) was used as the final step of the purification running into redox buffer at pH 8.6.

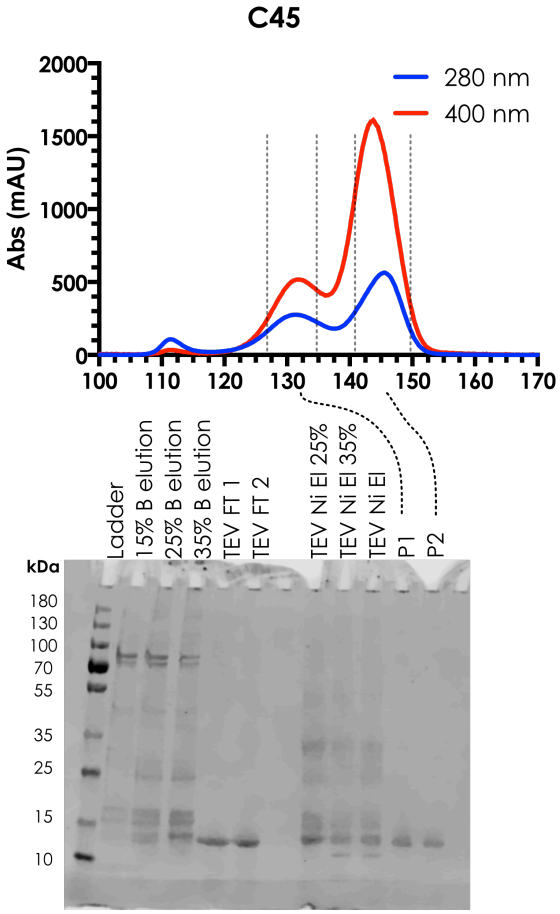


Figure 2.3.2 SEC and SDS-PAGE results for C45. *The 280 nm and 400 nm SEC trace is shown for C45. The x-axis shows the accumulated volume (in ml). The first peak is the void volume eluting around 43 ml after injection. Below the trace is the SDS-PAGE results of the purification. The lanes start with the ladder then 15, 25 and 35% elution buffer which is a wash of proteins weakly associated to the first nickel affinity column. The "TEV FT" 1 and 2 lanes are early and late in the flow through from the second nickel affinity column after cleavage of the his-tag. The "TEV Ni El" lanes show the elution from the second nickel affinity column, containing contaminants that bind to nickel and un-cleaved C45 sample protein, at 25 35 and 100% elution buffer. Lanes labelled as P1 and P2 are the two peaks on the SEC trace. Elution from the first nickel affinity column after the TEV elution" is the elution from the second nickel column. The remaining bands are fractions from the SEC purification of the flow-through from the second nickel column.*

The new maquette, C45, expressed and incorporated heme well. Figure 2.3.2 shows a size exclusion chromatography trace and SDS-PAGE of a purification of C45. The split peak seen in SEC was also seen in the parent maquette but it is clear from the SDS-PAGE that it is the same protein in both peaks.

2.3.3 UV-visible Spectroscopy

The UV-Visible spectra showed a change in the Soret λ_{max} value between the bis histidine-ligated parent protein C46 and the designed monohistidine-ligated C45. This suggests that the coordination state of the heme had changed. The ferrous (Fe^{2+}) spectrum for C45 exhibits a split Soret peak, indicating a mixed spin state population.

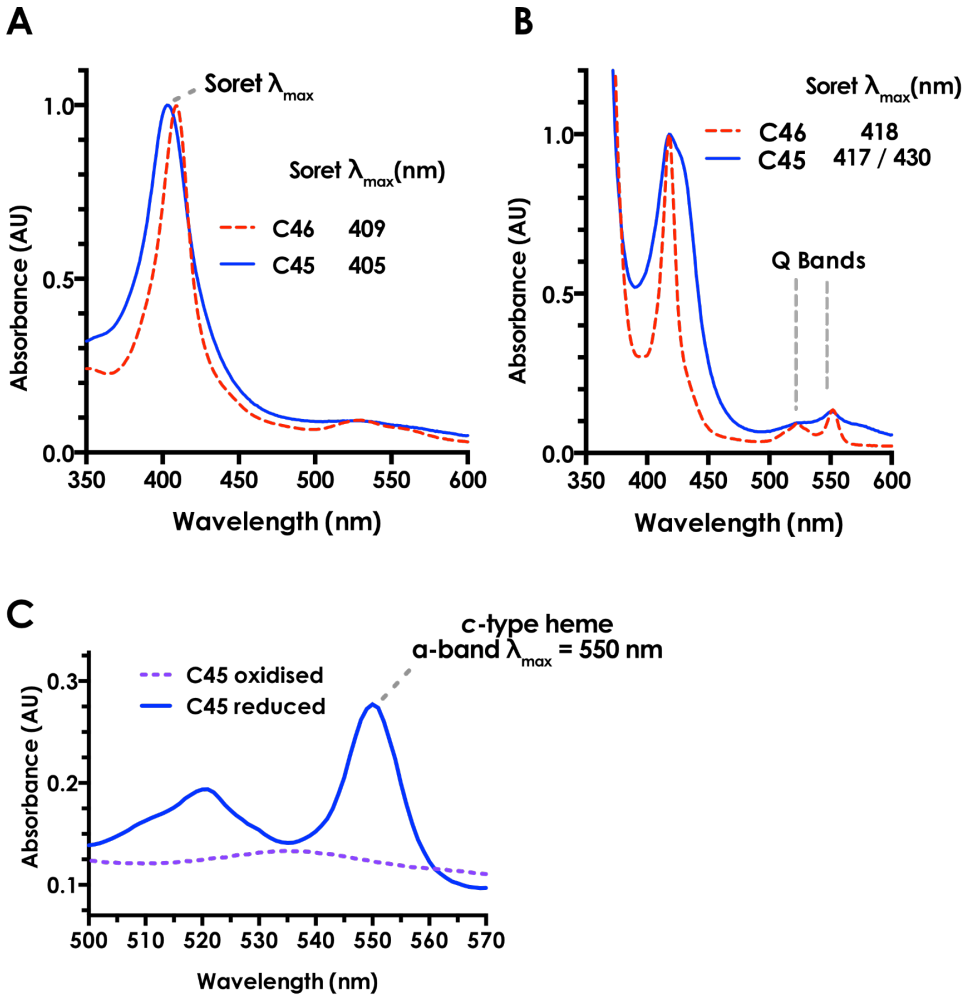


Figure 2.3.3 Maquette UV- visible spectra.

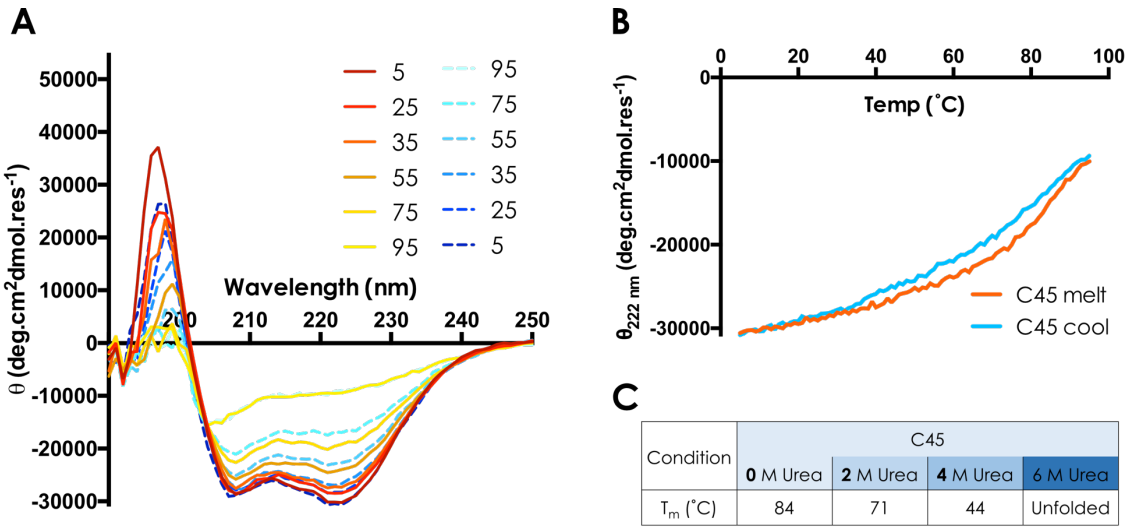
(A) *Shows a comparison of the ferric protein spectrum of C45 and C46.* ***(B)*** *Shows a comparison of the ferrous protein spectrum of C45 and C46.* ***(C)*** *Is the reduced and oxidised spectra of C45 during the Pyridine hemochrome assay showing the characteristic a-band absorbance at 550 nm*

C45 with imidazole added shows a similar peak to the bis-histidine ligated C46. The split peak of the ferrous spectrum of C45 is similar to the spectrum of Cytochrome *c'*. Cytochrome

c' is a histidine ligated *c*-type hemeprotein that has a water molecule bound to the distal heme face in the resting state. It exhibits Soret λ_{max} values of at 425 nm and 435 nm in similar buffer conditions [257]. The similarity Cytochrome *c'* and the changes observed on mix C45 with imidazole suggest that the distal site is now ligated with water. The pyridine hemochrome concentration assay showed a characteristic *c*-type heme peak at 550 nm the (Q-band α peak).

2.3.4 CD spectroscopy

The CD melt and cool showed that the activating mutation had not destabilized the protein. It is still extremely robust with a T_M=85 ˚C, similar to C46 (T_M = 84 ˚C). C45 refolds after denaturation to a similar helical content of the starting structure. The CD traces suggest that the protein is not completely denatured even at 95 ˚C as there is still some helical signal. The 222 nm melt and cool does not plateau indicating incomplete denaturation. Performing the same experiments with different concentrations of urea lowered the melting temperature of the protein (Figure 2.3.4.C and Figure 2.3.5). This highly helical signal demonstrates that the C45 has strongly defined secondary structure and is folded at room temperature.



C

Condition	C45			
	0 M Urea	**2** M Urea	**4** M Urea	6 M Urea
T_m (˚C)	84	71	44	Unfolded

Figure 2.3.4 CD results for C45. (A) *CD spectrum for C45. Scans of melts for 5-95 ˚C are displayed as solid lines going from red to yellow. Scans of cooling experiments from 95-5 ˚C are displayed as dashed lines from light to dark blue.* ***(B)*** *the mean residue ellipticity (θ) at 222 nm during both melt and cool. Table* ***C*** *gives the results for C45 with the T_m of unfolding at 3 urea concentrations. All experiments were conducted with 10 μM protein in Redox buffer at pH 8.6.*

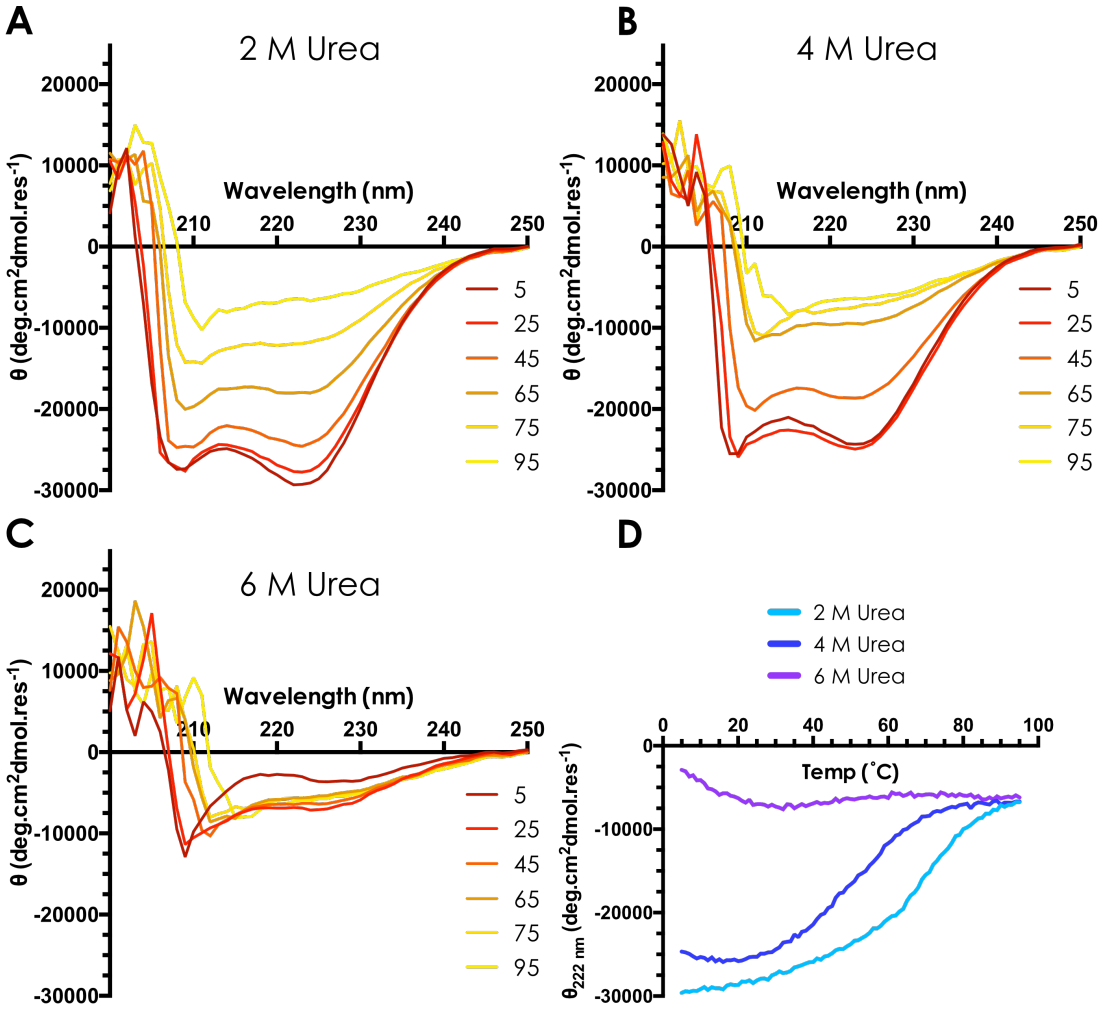


Figure 2.3.5 CD spectrum of C45 with urea. ***(A), (B)*** *and* ***(C)*** *show CD spectra at increasing temperature for C45 in 2, 4 and 6 M urea respectively.* ***(D)*** *shows the thermal melt trace at 222 nm for the three different urea concentrations.*

2.3.5 NMR

The high melting point observed in CD experiments encouraged further structural investigation. Regrettably, with many maquette proteins it is difficult to acquire atomistic structural information as they do not readily crystallise. The long and flexible loop regions are likely to be a key reason why the maquettes used in this project have not formed crystals. Maquettes are also flexible, dynamic and lack the design of rigorous core packing interactions seen in more rigid coiled-coil structures [46]. Both traits present a challenge to crystallisation. As a small soluble protein, NMR spectroscopy on a ^{15}N enriched sample could offer a means to determine high resolution structural information. In protein NMR, iron (III) causes paramagnetic broadening of the spectra. With this in mind, both Fe^{3+} heme and metal-free porphyrin-containing samples were produced. Spectra of C45 with and without iron can be seen in Figure 2.3.6.

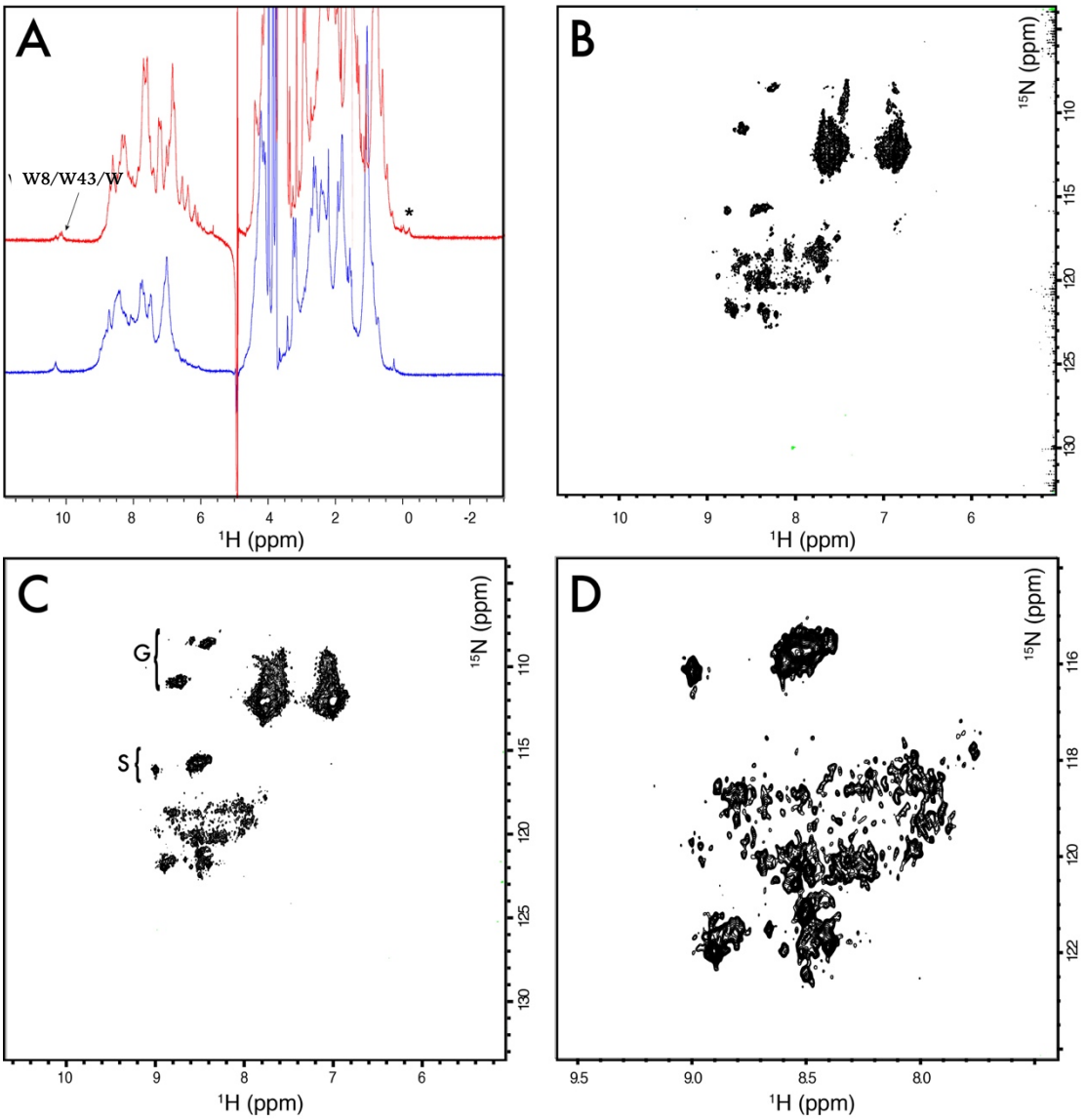


Figure 2.3.6 NMR of C45

*(**A**) 1D ^{1}H NMR spectra C45 with Fe^{3+} (blue) and demetalated C45 (red), * denotes up-field shifted methyl peaks. (**B, C**) ^{1}H-^{15}N SoFAST HMQC spectra of C45 (**C**) and demetalated C45 (**B**). (**D**) Expansion of (C) showing the numerous dispersed weak peaks.*

In the demetalated 1D ^{1}H NMR spectrum, three indole NH peaks are clearly resolved (corresponding to W8, W43 and W80). In addition, several up-field shifted methyl peaks are clearly visible indicative of core structuring. In the Fe^{3+} coordinated C45, only a single tryptophan resonance is visible. This may be cautiously assigned to either of the tryptophans separated from the paramagnetic broadening effects of the iron (W8 or W80) or over-lapped peaks for both. The helices of C45 have near identical sequences. This has contributed to the broadening of peaks in the HSQC experiments. Peaks corresponding to the random coil chemical shifts of glycine (GG and SG pairs) and serine found in the connecting loops are also

observed. A number of low intensity resonances in Figure 2.3.6.C, can be seen as a consequence of dynamic conformational heterogeneity and paramagnetic broadening from the Fe^{3+}. This is consistent with observations from other related maquette systems. C45 differs from intrinsically disordered or unfolded proteins showing numerous dispersed weak peaks (Figure 2.3.6.D) [66].

C45 has a highly repetitive sequence and a significant degree of conformational flexibility, which has limited the structural information that can be gained from NMR and attempts at crystallographic studies. These results, in combination with CD spectroscopy, show C45 to be similar to other reported maquette proteins. C45 clearly has defined secondary structure, randomly coiled loops and some core packing.

2.3.6 Molecular Dynamics

Molecular dynamics (MD) simulation is a powerful tool used for assessing protein dynamics. For easy compatibility with other software and to make use of GPU acceleration, it was necessary to compare running MD simulations with both GROMACS and AMBER16 using the CHARMM27 force field. Both AMBER and GROMACS simulations were completed successfully. A comparison of the final structures and root-mean-square deviation (RMSD), from the starting structure, throughout the trajectory is displayed in Figure 2.3.7. The average structural identity during the simulation was very high, $Q_H = 0.8347$ (SD = 0.0367). A similar value could be expected when comparing different frames of the same simulation. A Q_H of 1 is expected for identical structures. Both simulation methods are in agreement, which is unsurprising since both programmes use the same force field.

The starting atomic coordinates for the MD were generated computationally, partly using a homology based on a related maquette X-ray structure. The model has been used previously and was modified to match the sequence of C45 [69]. The MD results show a stable RMSD over 100 ns around 1.5 Å. Generally, this is considered to indicate that the model is in an energy minimum and is a value you would expect when simulating an experimentally derived structure of a natural protein.

Four additional repetitions of the AMBER MD simulations were run for C45. The MD results for C45 show a stable four-helical bundle with dynamic loop regions. The helices show little structural perturbations. There are two kinks in the α-helical structure, one on the fourth helix at the heme binding site and one on the second helix on the other side of the bundle.

The well-defined helices and dynamic loop regions are similar to natural four-helix bundles like cytochrome b_{562} [50].

The heme propionates face into the solvent and much of the γ and δ-heme edge are also solvent accessible. The bulk of the heme is buried in the hydrophobic core of the bundle. The site of the distal histidine ligation in the parent maquette, C46 H46, is now a phenylalanine (F46), which contributes further to the hydrophobic nature of the distal heme site. The MD simulations show that F46 in C45 is in a position to form π-stacking interaction with the heme. In natural heme proteins π-stacking interactions have been seen to be important in stabilising ligand binding as well as modulating the redox potential of the heme [118].

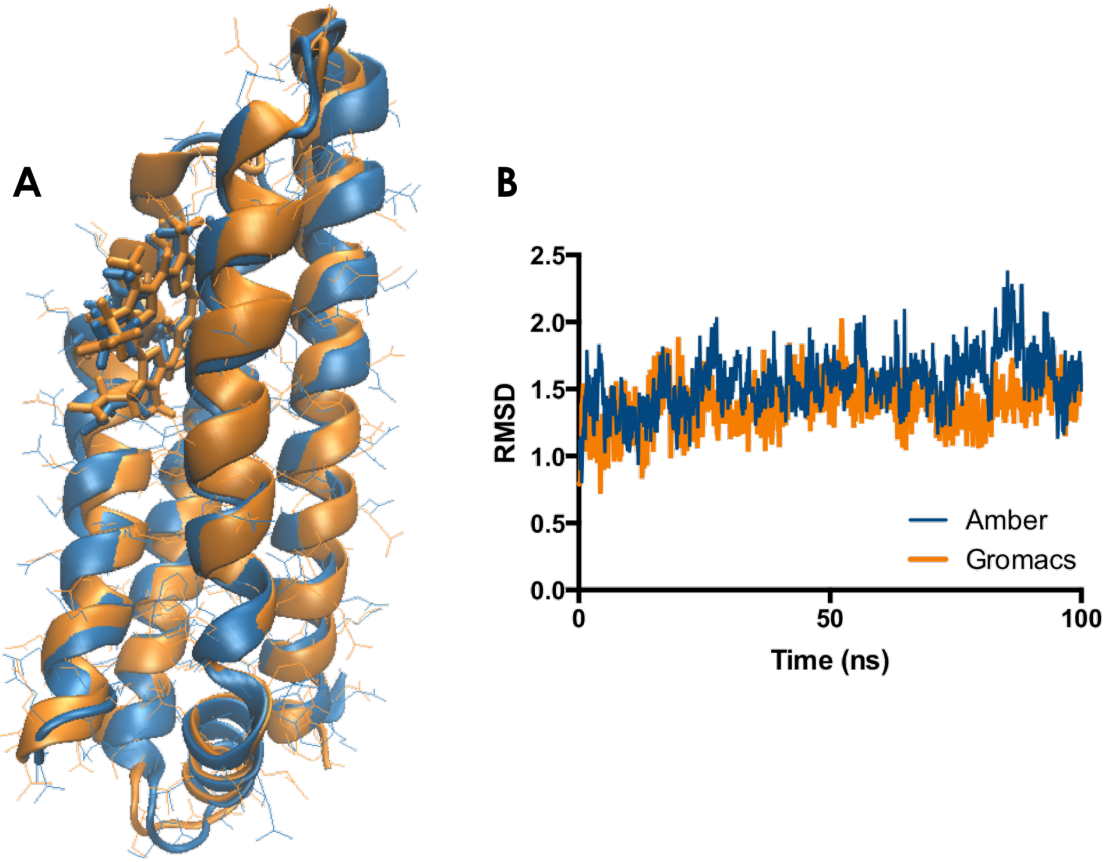


Figure 2.3.7 C45 MD comparison.
(A) *Is a comparison between the last frame of the GROMACS (Orange) and AMBER (Blue) simulations, this is after 100 ns of simulation.* ***(B)*** *Shows a comparison of AMBER and GROMACS RMSD over the course of the simulation. Each RMSD is calculated against its first frame using only backbone atoms. The Average values of the RMSD for AMBER was 1.550 (SD = 0.2177) and for GROMACS was 1.387 (SD = 0.1832).*

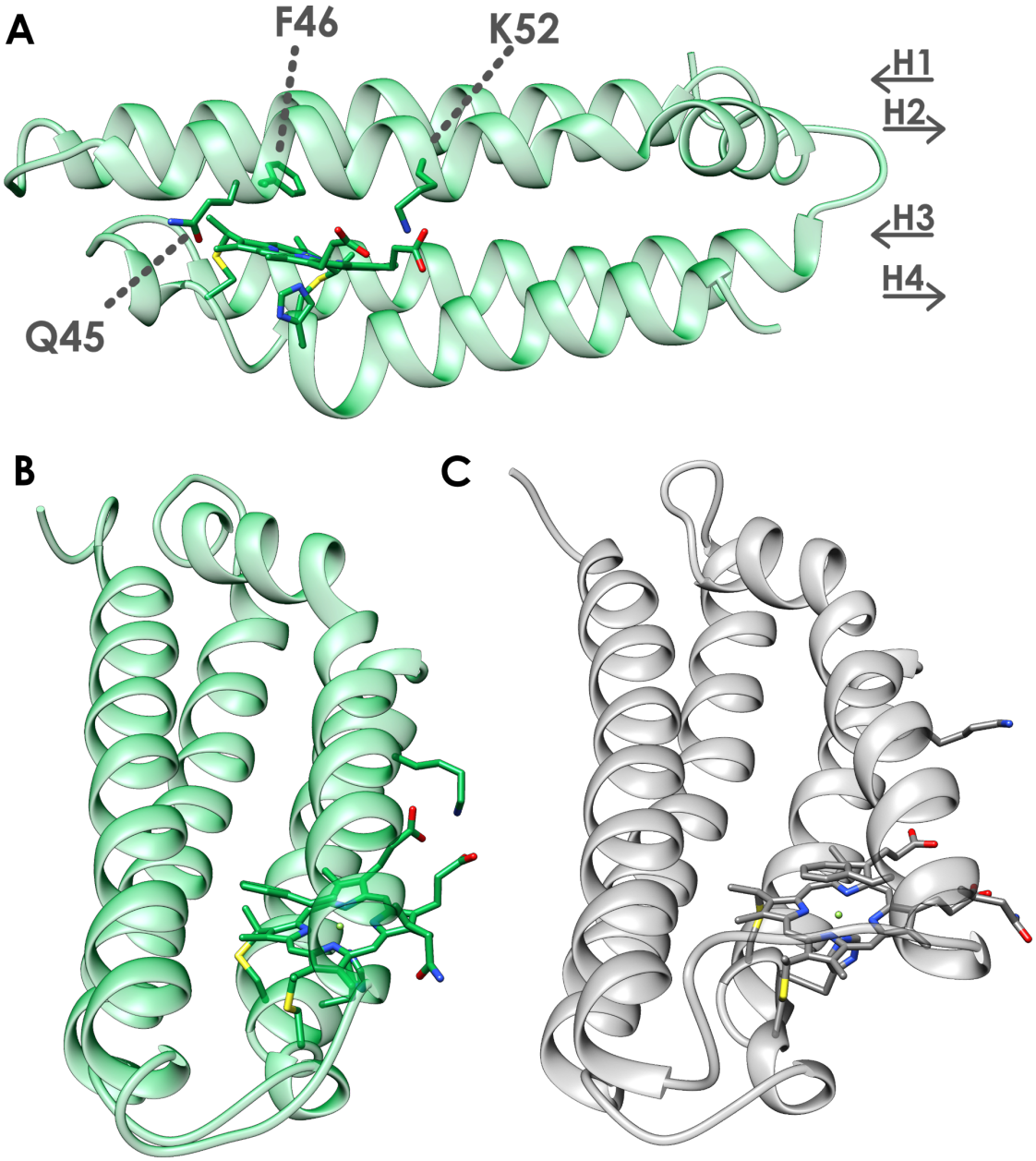


Figure 2.3.8 MD structures of C45. ***(A)*** *and* ***(B)*** *are images of the same frame from MD simulation of C45 for two different viewpoints. In (A) residues Q45, F46 and K52 and the direction of the helices are labelled.* ***(C)*** *is from the MD simulation where a large movement of the second helix occurs. Hydrogen atoms are present but not depicted.*

Q45 and K52 are polar residues that can closely associate with the heme during MD and could play a role in stabilising charge during the reactive cycle (labelled in Figure 2.3.8.A). Q45 can be seen as close as 5.8 Å to the heme iron. K52 forms hydrogen bonds with the heme propionates, with the ammonium nitrogen around 6 Å from the heme edge. The pH dependence of C45 ‘s spectroscopic and kinetic profile suggest a residue with a pK_a of 9.1 is involved in the reactive cycle. Lysine is the most likely candidate considering its proximity to the heme.

In one of the repeated MD runs there is a notable shift of the second helix, directly above the heme. This was only seen once and seemed to be caused by movement in the loop region causing a shift in the helix. Figure 2.3.8.C shows the cavity that opens up above the heme as a result of this movement. The helix movement occurred after 60 ns and persisted to 100 ns.

2.3.7 Preliminary computational folding with Rosetta

With new, timely advances in protein structure prediction, an attempt was made to fold C45 with the Rosetta *ab initio* folding software. Baker *et al.* have constantly demonstrated that Rosetta is extremely successful at predicting the structure of many natural proteins even before an experimental structure is made public [30, 258, 259]. It achieves this success using primary amino acid sequences alongside small structural fragments from the PDB. At the time of writing, these predictions seldom include cofactors. Regarding examples that do implicitly include cofactors, modelled cofactors often become bespoke descriptions of individual systems rather than general solutions. Artfully creating Rosetta compatible parameters for *c*-type heme was outside the scope of this project and, to date, there are no published examples of *ab initio*-folded heme proteins to provide hope that it is an endeavour worth pursuing. With this in mind, it is necessary to state that the results discussed here are very much preliminary.

Despite some considerable effort, all simulations could only run successfully with the histidine rotamer singly protonated on Nε nitrogen (HIE not HID in the AMBER naming convention). This shortcoming meant that all simulations with the heme contain a steric clash between the proximal histidine ligand and the heme iron. This clash, along with other restraints imposed to bind the heme in the correct orientation, have resulted in less negative Rosetta scores for simulation with heme than without. It was not possible to calculate a comparable Rosetta score for the existing C45 MD structure because of the implementation of the constrains to bind the heme. The results for both the simulation with and without the heme do not appear to have converged on a single low energy structure. For a converged simulation, the plots of the Rosetta score against the RMSD (of the structure compared to the native or lowest energy structure) should form a funnel converging on low energy and low RMSD structures. The plots in the bottom right of Figure 2.3.9 and Figure 2.3.10 should look more like this funnel with the lower scores sampling a reduced number of more similar conformations.

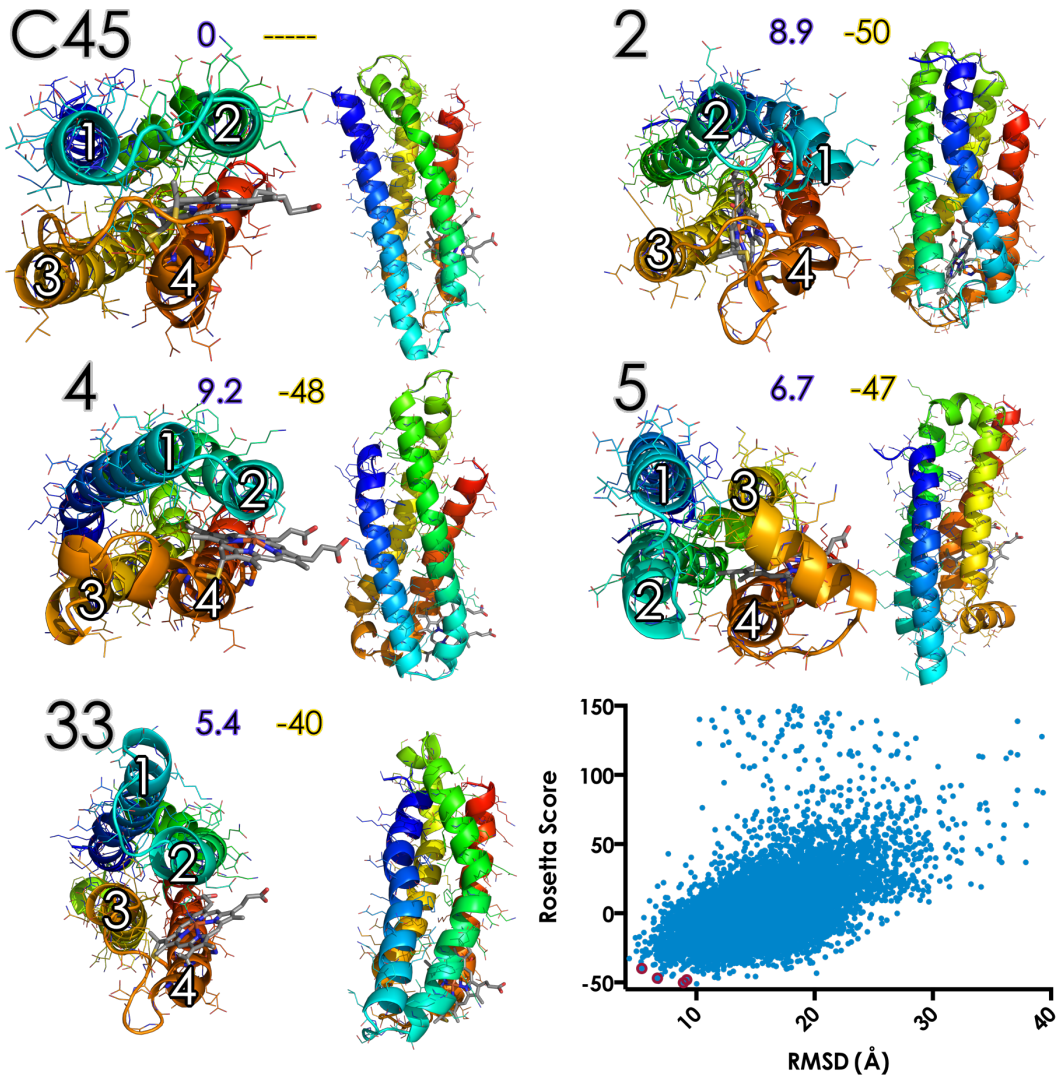


Figure 2.3.9 Rosetta structures with heme.

The first structure (top left) is a MD snapshot of the existing model of C45. The subsequent structures are the results of the Rosetta ab initio calculation they are named by their Rosetta score, ranked 1 to 10,000. The second number following the black name/rank for each structure is the RMSD (Å) compared to the existing MD model of C45. The third number is the Rosetta score for that structure. The structures are coloured by residue number, blue for the first to red for the last. Heme is in grey. The white numbers on the structures are the helix number. The plot in the bottom left is a comparison of all the Rosetta results, red points denoting the structures displayed in this figure.

The folding simulations **with heme** resulted in less realistic structures. Two of the top ten lowest energy structures did not manage to fold into a four-helix bundle, as one helix was extended away from the bundle in these cases. The top scoring results could be grouped into three main folded topologies (not including extended chains). The first is shown in Figure 2.3.9.2 where the helices are all linked by "small" loops. The longer loop does not reach diagonally across the bundle. The heme propionates in this orientation face directly into the

bundle, which seems unlikely a native-like structure. The second is shown in Figure 2.3.9.4 & 33 and exhibits the same overall topology as the existing C45 model. The longer loop links helix 2 and 3 diagonally across the bundle with the heme propionates facing out to the solvent. The third topology is shown in Figure 2.3.9.5. In this model, helix 2 and 3 are swapped relative to each other, the longer loop still linking them diagonally across the bundle, with the heme propionates facing out into the solvent.

The folding simulations **without heme** resulted in structures with the same three topologies as the simulations with heme. The first in Figure 2.3.10-1, without the diagonal long loop across the bundle, comprised 7 of the top 10 lowest energy structures. Both the second and third helix topologies with the long diagonal loop were also seen in the top 10 lowest energy results. Figure 2.3.10.2 and 15 are similar to the existing C45 model. Figure 2.3.10-6 shows the third topology with helix 2 and 3 swapped. The fourth helix of all obtained structures would need to move considerably to accommodate a heme group.

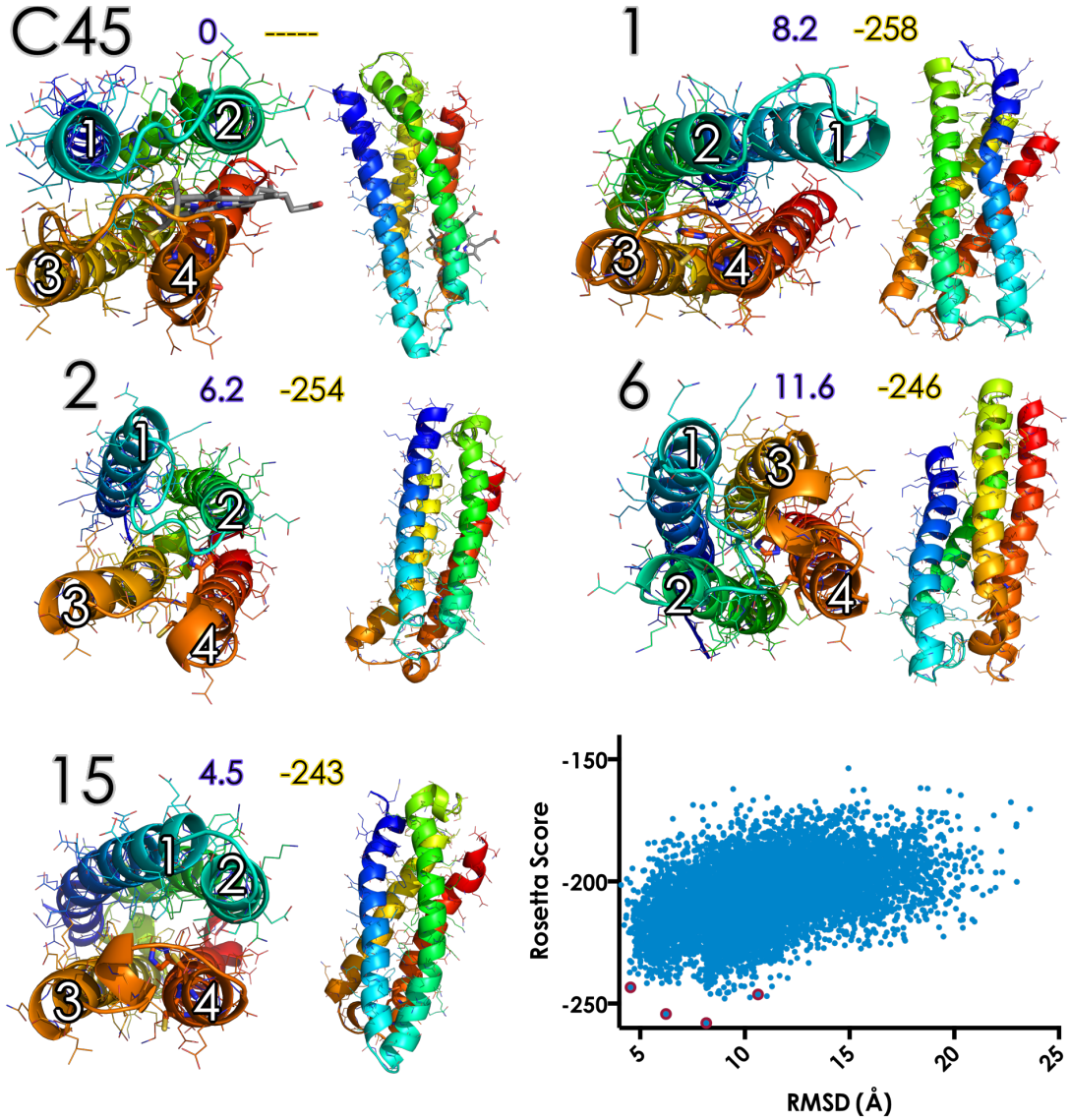


Figure 2.3.10 Rosetta structures without heme.

The first structure (top left) is a MD snapshot of the existing model of C45. The subsequent structures are the results of the Rosetta ab initio calculation. They are named by their Rosetta score, ranked 1 to 10,000. The second number following the black name/rank for each structure is the RMSD (Å) compared to the existing MD model of C45. The third number is the Rosetta score for that structure. The structures are coloured by residue number, blue for the first to red for the last. The white numbers on the structures are the helix number. The plot in the bottom left is a comparison of all the Rosetta results with the red points denoting the structures displayed in this figure.

2.3.8 Kinetic assays for Ping-pong fitting

To test the peroxidase activity of C45 or C46, they were simply mixed with an excess of hydrogen peroxide (1 mM H_2O_2, 0.5 μM C45 or C46) at room temperature in a small volume cuvette. The reaction was observed using UV-visible spectroscopy, monitoring the Soret absorbance band corresponding to the most intense heme π-π* transition. There was no significant change in the UV-visible spectra of the bis-histidine ligated maquette, C46, on the addition of peroxide. Conversely, with C45 there is a red shift of the Soret λ_{max} (406 to 411 nm) and bleaching of the heme signal. This bleaching corresponds to oxidative damage of the heme and is likely a result of generating heme reactive species (compound I and II) in the absence of an electron donating substrate. Compound I and II are highly redox active and capable of oxidising protein or heme, leading to the degradation of the Soret absorbance.

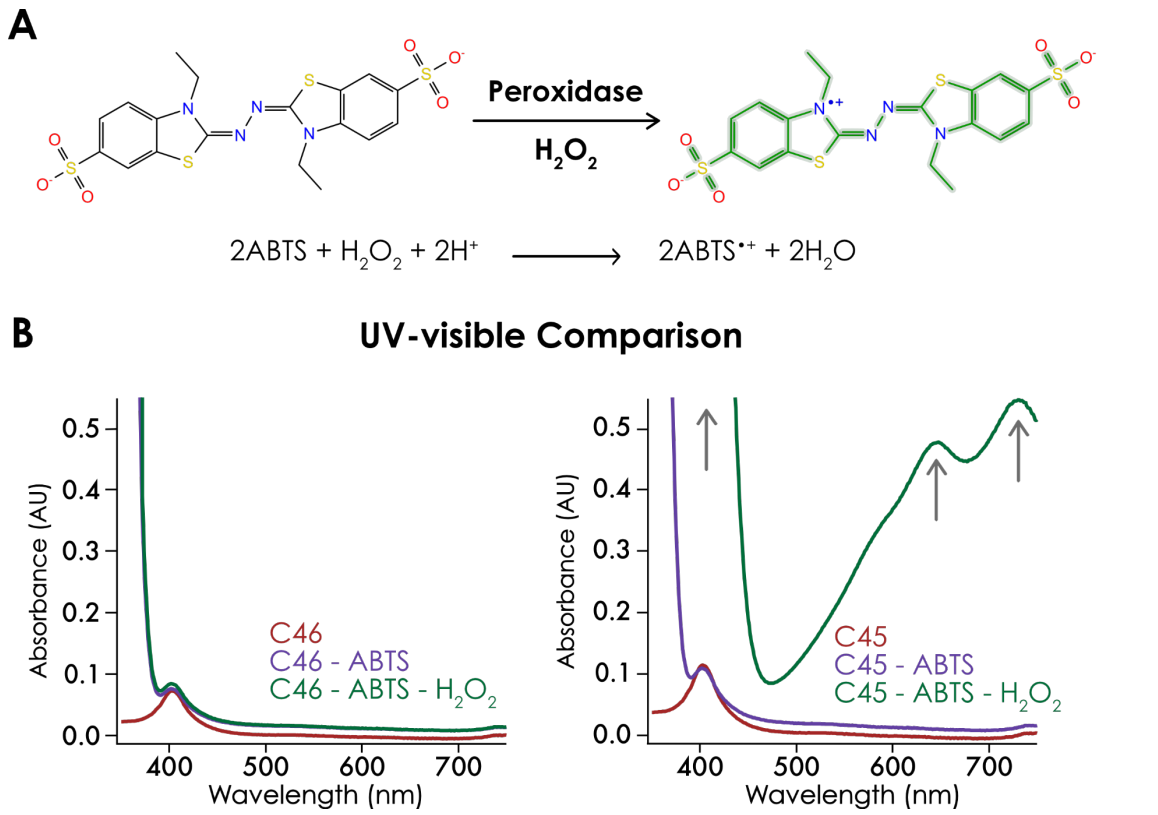


Figure 2.3.11 ABTS reaction with C45

*Panel **(A)** Shows the structure of ABTS and Radical cation in green, with equation below. **(B)** Shows the visible absorbance changes on addition of H_2O_2 to C46 (left) and C45 (right). The **red** line is 0.1 μM protein. The **purple** line is Protein with ABTS (1 mM). The **green** line is 120 seconds after the addition of 1 mM H_2O_2 to 1 mM ABTS and 0.1 μM protein*

ABTS, a widely-used electron donating peroxidase substrate, was used to investigate the peroxidase activity of C45. ABTS oxidation generates an intensely coloured $ABTS^{\bullet+}$ radical cation allowing the reaction to be monitored using UV-visible absorption spectroscopy. In the absence of a catalyst, solutions of H_2O_2 mixed with ABTS exhibit no spectroscopic changes over the course of the experiments at room temperature and pH 8.6. Figure 2.3.11.B shows the spectral changes on the addition of hydrogen peroxide to C46 and C45. C46 shows almost no peroxidase-type activity, whereas C45 efficiently oxidises ABTS to the corresponding ABTS radical cation.

The reaction of ABTS with C45/H_2O_2 can be used to study the kinetic mechanism of this *de novo* peroxidase. Steady-state assays were conducted by individually varying the concentration of both substrates (ABTS and H_2O_2) at the optimum pH (see Section 2.3.9) and recording the initial rates for ABTS oxidation. Taking a phenomenological approach, the data fit best to a standard ping-pong mechanism. The general form of the ping-pong mechanism is seen in scheme in Figure 2.3.12.A . It involves the modification of the enzyme (E) by substrate A forming E' which is then able to turn over a second substrate B. In the standard mechanism A and B are not bound in the enzyme at the same time.

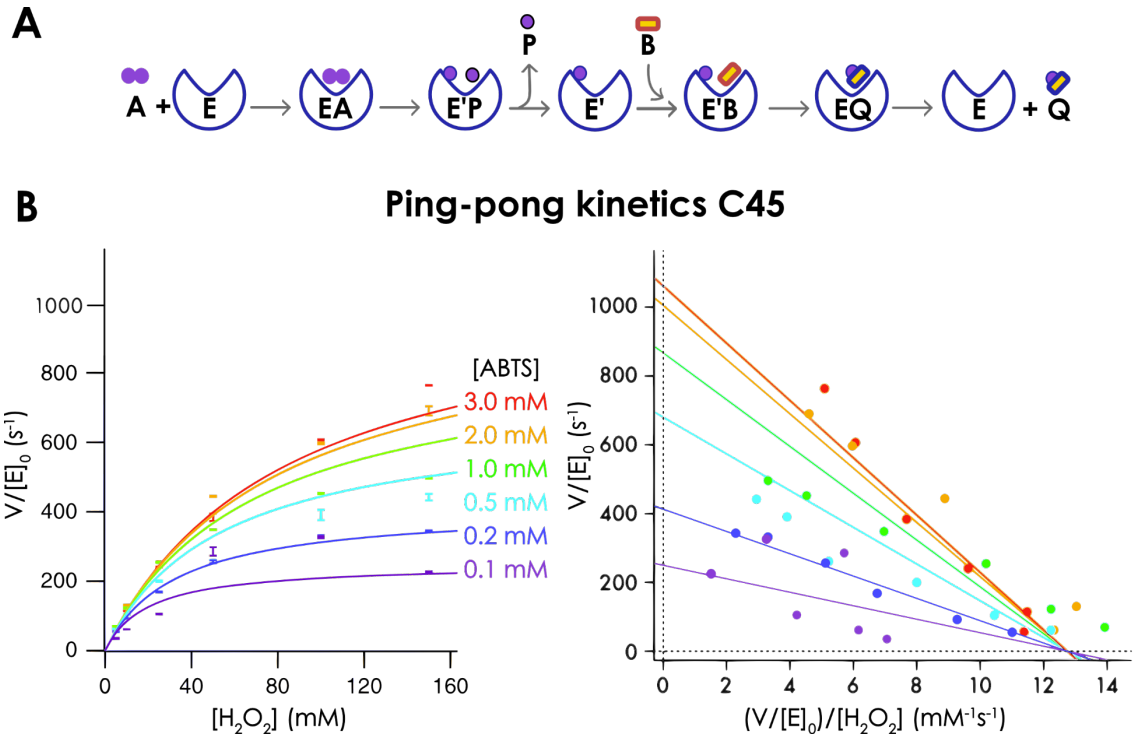


Figure 2.3.12 Ping-pong kinetics C45

(A) *Shows a general scheme of the ping-pong mechanism*

(B) *Shows the result of ping-pong fitting on the steady-state kinetics of various concentrations of ABTS and H_2O_2. On the right is an Eadie-Hofstee plot of the result.*

The fit results for C45:

$$k_{cat} = 1200\ s$$

$$K_{m(ABTS)} = 379\ \mu M \qquad K_{m(H2O2)} = 94\ mM$$

$$k_{cat}/K_{m(ABTS)} = 3.2 \times 10^{6}\ Ms \qquad k_{cat}/K_{m(H2O2)} = 1.3 \times 10^{4}\ Ms$$

Equation 2.4 is the equation for the ping-pong mechanism that can be derived using the King-Altman Method and was subsequently used to perform the fitting [260]. Under the steady-state approximation, equation 2.4 describes the rate (v) using the Michaelis–Menten constants $K_{m(A)}$ and $K_{m(B)}$ for substrate A and B respectively. The results of the fitting for C45, displayed above, boasts the best yet recorded catalytic efficiency (k_{cat}/K_M) for a *de novo* enzyme.

$$v = [\mathrm{E}]_0 k_{cat} \frac{[A]_0[B]_0}{K_{M(B)}[A]_0 + K_{M(A)}[B]_0 + [A]_0[B]_0} \tag{2.4}$$

2.3.9 pH Dependence

The pH-dependence of ABTS oxidation by C45/H_2O_2 was explored whilst varying the pH between 6 and 11. The results of the pH dependence assays can be seen in Figure 2.3.13. The optimum pH for catalytic activity for C45 was determined by plotting $k_{cat(ABTS)}$* of the ABTS oxidation against pH. The data was fitted to a double pH curve (equation 2.5). The Optimum pH was found to be 8.6 for ABTS. The reactions for the pH dependence of H_2O_2 were performed using a stopped-flow spectrophotometer and required high concentrations of ABTS. At high pH there appeared to be a problem with obtaining reliable and statically relevant results, possibly due to stability problems of high concentrations of ABTS at these pH values.

The two pK_a values from the double pH curve fitting (pK_{a1} and pK_{a2} in equation 2.5) were calculated to be 8.0 and 9.1. The lower pK_a is likely to arise from the deprotonation of H_2O_2 in the formation of compound 0. The second pK_a is expected to be the deprotonation of a residue that participates in the reactive cycle in some way. From the MD simulations, this is most likely to be K52.

$$f(x) = \frac{\left(k_0.10^{(pKa1\,-\,x)}\right) + k_1 + \left(k_2.10^{(x\,-\,pKa2)}\right)}{1 + 10^{(pKa1\,-\,x)} + 10^{(x\,-\,pKa2)}} \tag{2.5}$$

pH	k_{cat}* (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
6	0.34 ± 0.02	1.81 ± 0.66	1.88± 0.71 x10^5
7	0.47 ± 0.03	0.41 ± 0.23	1.15± 0.65 x10^6
8	0.91 ± 0.07	1.8 ± 0.69	5.06± 1.98 x10^5
8.5	1.11 ± 0.06	2.22 ± 0.61	5.00± 1.40 x10^5
9	1.05 ± 0.13	7.14 ± 3.18	1.45± 0.67 x10^5
10	0.55 ± 0.02	0.72 ± 0.16	0.76± 1.74 x10^5
11	0.53 ± 0.05	4.07 ± 1.42	1.30± 0.46 x10^5

Table 2.3.1 pH dependence Michaelis–Menten fit values. *The concentration of ABTS was varied with C45 and H_2O_2 and the initial rates fitted to Michaelis–Menten curve. The k_{cat}* is because reactions were conducted at rate limiting H_2O_2 concentrations (100 μM). The ± error are the standard deviation.*

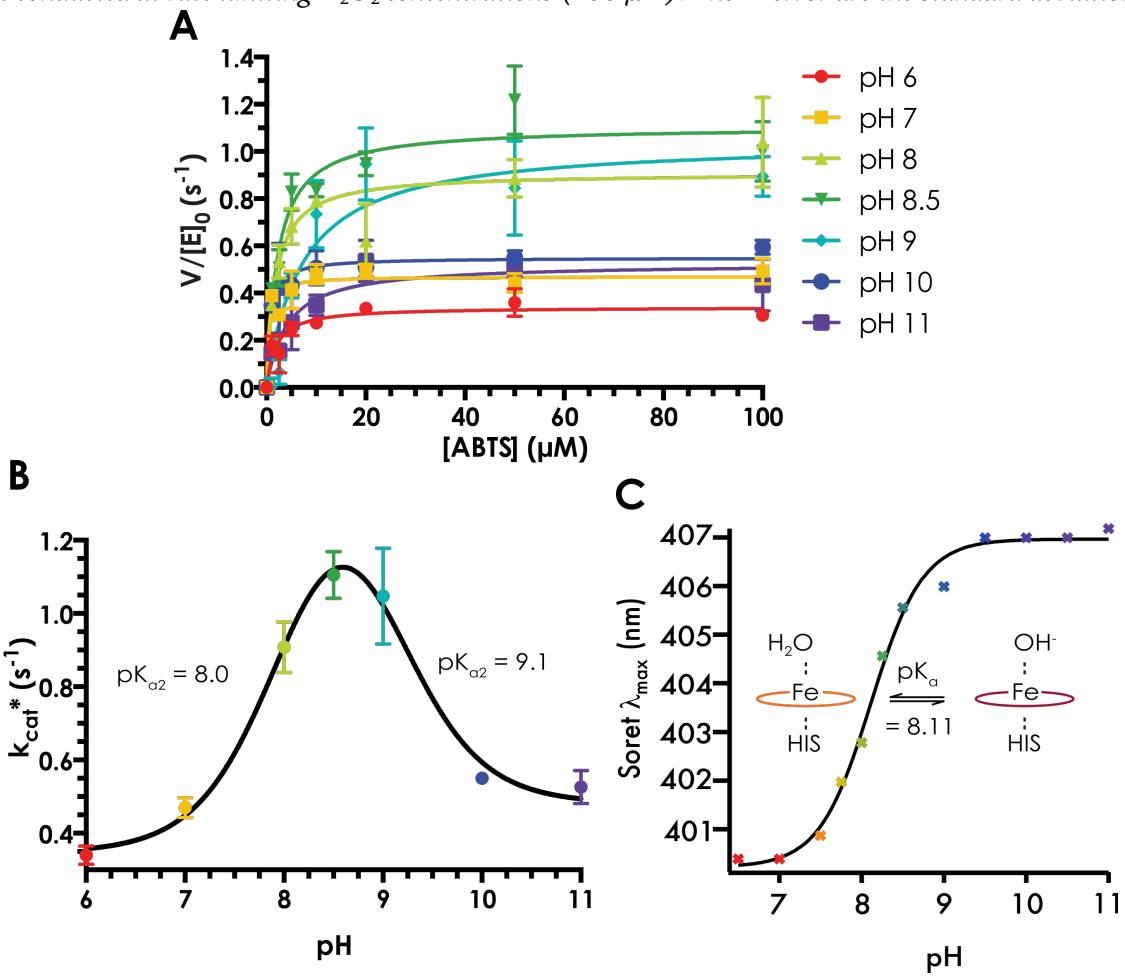


Figure 2.3.13 pH dependent steady-state kinetics. ***(A)*** *shows the Michaelis–Menten results for reactions with C45 H_2O_2 and various concentration of ABTS at different pH. (B) show the pH dependence of the k_{cat}* fitted to a double pH curve using equation 2.5. (C) is the pH dependent changes to the Soret λ_{max}. The data for (C) was collected by Daniel Watkins [70]. The error bars show the standard deviation.*

pH	E_M1	E_M2
6.0	-166.2 ± 3.5	-107.3 ± 3.2
7.5	-187.2 ± 6.5	-126.2 ± 5.9
8.6	-173.1 ± 0.4	NA
10.0	-208.3 ± 1.3	NA

Table 2.3.2 C45 *pH dependent midpoint potentials (E_M). Values are shown in mV vs NHE.* *OTTLE results have been fit to are displayed for C45 at four pH values. This data was collected by Daniel Watkins [70]. The error bars show the standard deviation.*

2.3.10 Temperature-Dependent kinetics

The high thermal melting temperature observed in CD experiments made it clear that C45 is structurally tolerant to temperature. To test the effect increasing the temperate had on the activity of C45 its kinetic assay with ABTS and H_2O_2 were performed at various temperatures. Increasing the temperature increased both the k_{cat}* and the K_M. The maximum k_{cat}* and K_M were recorded at 70 ˚C (seen in Table 2.3.3). The data for this experiment was collected by Daniel Watkins [70].

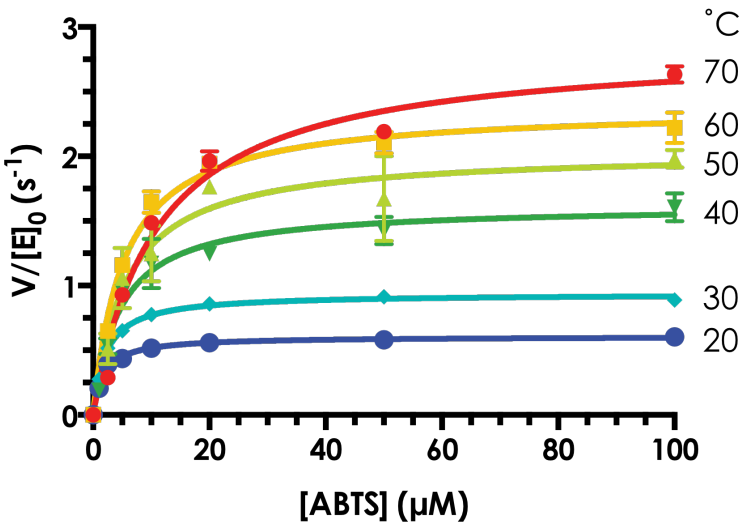


Figure 2.3.14 Temperature-Dependent steady-state kinetics. *Reactions were performed with ABTS, 0.25 µM C45 and 100 µM H_2O_2. Fit values from the initial rates are shown in Table 2.3.3. Error bars are the standard deviation.*

Temperature (˚C)	k_{cat}* (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
20	0.61 ± 0.01	1.72 ± 0.15	3.53 ± 0.31 x10^5
30	0.94 ± 0.01	2.16 ± 0.13	4.35 ± 0.27 x10^5
40	1.62 ± 0.06	4.55 ± 0.72	3.56 ± 0.58 x10^5
50	2.04 ± 0.11	5.57 ± 1.22	3.66 ± 0.83 x10^5
60	2.38 ± 0.07	5.26 ± 0.58	4.52 ± 0.51 x10^5
70	2.86 ± 0.17	10.8 ± 2.09	2.65 ± 0.54 x10^5

Table 2.3.3 Temperature dependence Michaelis–Menten fit values. *The concentration of ABTS was varied with C45 and H_2O_2 and the initial rates fitted to Michaelis–Menten curve. The k_{cat}* is because reactions were conducted at rate limiting H_2O_2 concentrations (100 μM). The ± error are the standard deviation.*

2.3.11 Computational binding screen

A computational binding screen was performed on MD snapshots of C45 and 19 small molecule peroxidase substrates. The Bristol University Docking Engine (BUDE) was used to search for low energy binding conformations. Each generated position of a ligand on the surface of the protein is termed a pose. The top 50 binding poses for each ligand were then tested in a 5 ns MD simulation. For TCP, only the top 40 poses were selected, as all top 50 were in the same locus. This resulted in 940 (18 x 50 + 40) starting structures for simulation. 18 starting structures produced badly minimized simulations that crashed during the production MD and were not investigated further. The 922 completed 5 ns MD trajectories for the 19 ligands represent approximately 77 days of GPU time.

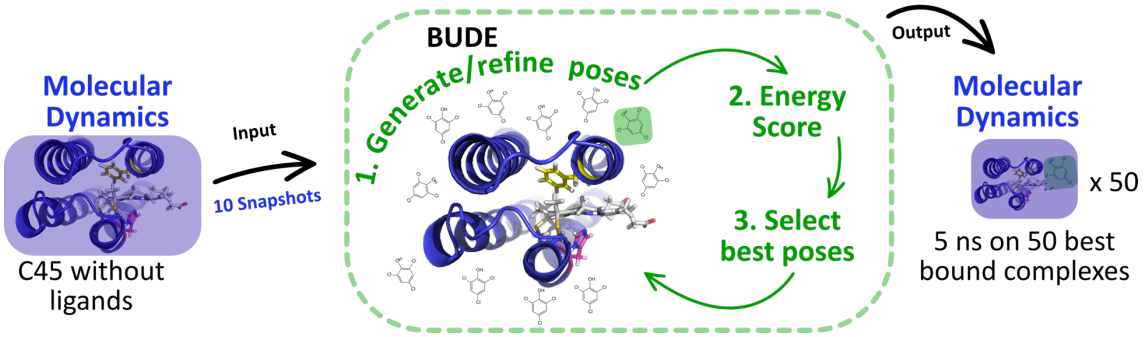


Figure 2.3.15 BUDE process illustration. *Ten MD structures were used as the input into the BUDE calculation. The green dashed box represents the BUDE calculation which iterates through millions of binding poses using a genetic search algorithm. Each generation is made up of a cycle of generating or refining the binding poses, scoring and then selecting the highest energy poses to be used in the next cycle. The top 50 binding energies from the BUDE calculation were then tested in a 5 ns MD simulation.*

The results from the computational binding screen reveal several proposed binding hot spots close to the heme. To probe the interaction of substrates with C45 further and to discern which ligands share similar binding loci, a 3D clustering analysis was performed. Figure 2.3.16 shows a centroid representation of ligands grouped by the results of the clustering analysis at the end of the 5 ns trajectory. The pivot table in Figure 2.3.16 catalogues the binding hot-spots for each ligand.

Successful ligand poses in this context are counted as those that are still associated with the protein at the last frame of the 5 ns MD simulation. An association with the protein was defined as the heavy atoms of the ligand within 5 Å of any part of the protein or heme. The full structures of ligands at the end of the 5 ns production MD are displayed in Figure 2.3.18 and Figure 2.3.19.

Although somewhat extended, **cluster 3** (Figure 2.3.16) is the most frequent and widely sampled location with 52% of successful ligands bound in that site after MD. It is above the heme between helix 1 and 2 of the protein. The small phenolic substrates (TCP, TBP, Guaiacol), *o*-Phenylenediamine and 1,4-Dimethoxybenzene have a strong preference for binding on the side of cluster 3 closest to heme. This population within group 3, that is situated close to the heme, could be defined as the small phenolic binding pocket (SP pocket), although non-phenolic small molecules also bind. The majority of Luminol, *o*-Dianisidine, *p*-Anisidine and Isoniazid binding poses are also found in the SP pocket. The centre of mass of ligands in this position is around 10 Å from the heme iron, or around 7 Å from substrate to heme edge. These small non-charged substrates sit deep in the protein, in a location centered over leucine 22, shown in red inside the blue box in Figure 2.3.19. The distance between heavy atoms of ligands and the two closest tryptophan residues is approximately 11 Å (W43) and 22 Å (W8). 287 of 353 poses in the SP pocket before MD were found in cluster 3 after MD. Luminol, Isoniazid and 1,4-Dimethoxybenzene were more likely to migrate into **cluster 12** during the simulation. 27 luminol poses found in the SP pocket before the MD had dissociated after MD, indicating that luminol may bind less stably than TCP, where all 40 poses persisted without moving. Veratrole alcohol and Ascorbate have been categorized by the analysis as within cluster 3 but are closer to cluster 12, and are generally around 20 Å from the heme iron.

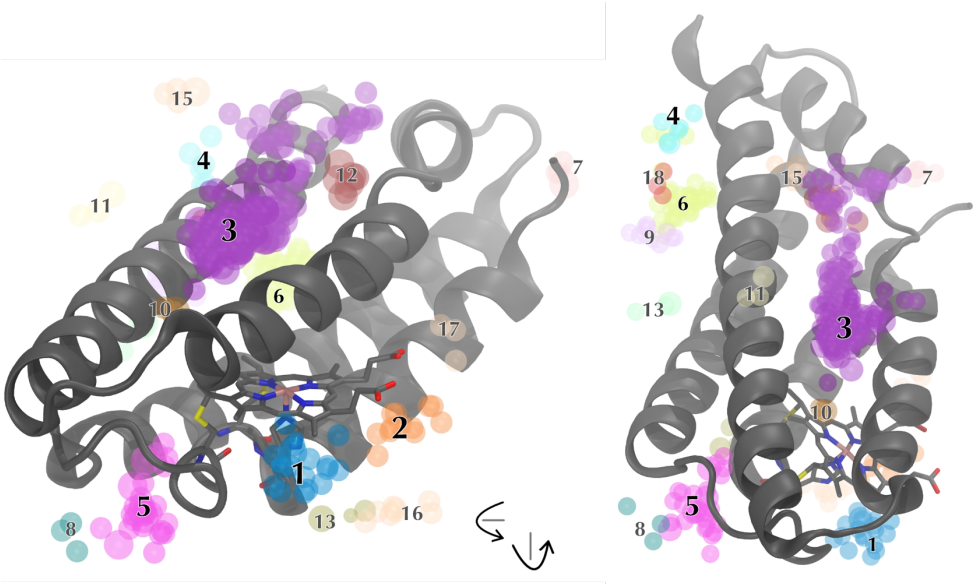


Figure 2.3.16 Binding Screen Clusters

Results from 3-dimensional cluster analysis of the last frame of 5 ns MD. Each dot represents the centre of mass of a ligand. The number assigned to each cluster is arbitrary. Ligands that were associated with the protein but not assigned a cluster are not displayed and given cluster number -1.

Below is a pivot table comparing ligands and assigned clusters. Ligands associated with protein but not clustered are given cluster number -1. Taking p-Anisidine as an example, 44 top binding poses were associated with to the protein at the end of 5 ns in the cluster 3 locus, 3 in other groups, 1 associated but not grouped and 2 dissociated from the protein.

Ligand	3	6	5	1	2	9	16	12	4	15	7	18	8	13	14	17	11	10	-1	Grand Total
p-Anisidine	44							1	1								1		1	48
Amplex Red	5	14	8	4		1			2		2		3			1			5	45
TBP	33	1	6											1					4	45
1,4-Dimethoxybenzene	35		2							1								1	2	41
Dihydroxyphenylalanine	2	10	1	4	2		8									2			12	41
Veratrole alcohol	19		5	1				5			1			1					8	40
TCP	40																			40
Reactive blue 4					8	7						3							21	39
o-Dianisidine	19	4				1		1	1	1							2		9	38
Isoniazid	30									1								1	5	37
Guaiacol	21		8	1														1	5	36
ABTS	2			5	3														23	33
o-Phenylenediamine	28														1				2	31
Tetramethylbenzidine	7	8		1		2			2					1					9	30
Ascorbate	16									1	1				1				5	24
Luminol	17																		5	22
5-Aminosalicylic Acid				1															8	9
Homovanillic Acid				2											1				4	7
4-Aminobenzoic acid																			3	3
Grand Total	**318**	**37**	**30**	**19**	**13**	**11**	**8**	**7**	**6**	**4**	**4**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**131**	**609**

The second most frequently sampled binding pocket is **cluster 6**. It makes up 6.1 % of successful poses and is situated on the opposite side of the protein to the heme between helix 1 and 3 of the model (Figure 2.3.16). The centres of mass of the ligands in cluster 6 are on average 27.1 Å from the heme iron or about 22 Å from the heme edge. Electron transfer at this distance is not impossible, but would be expected to be much slower [83]. This cluster is very close to the two tryptophan residues (<10 Å) at the further end of the maquette from the heme (W8 and 80).

Clusters 5 and **8** make up 5.4 % of successful hits and are situated in a cavity between helices 3 and 4 (Figure 2.3.16). Here TPB, Guaiacol, Amplex Red and Veratrole alcohol associate stably after 5 ns of MD. The centre of mass of ligands in this group are on average 11.9 Å away from the heme iron or around 6 Å from substrate to the heme edge. The distance between heavy atoms of ligands and the closest tryptophan residue is approximately 11 Å (W43).

Clusters 1 and **2** are in a similar position at the heme edge or interacting with the distal histidine. Combined, these groups make up 5.2 % of successful poses. Cluster 1 and 2 are found on the δ and γ heme edges respectively. The red box within Figure 2.3.19 is a heat map of heme atoms and the frequency of close contacts with ligands displayed. This analysis indicates that ligands that interact with heme do so mainly on the δ and γ heme edges. In general, the larger ligands displayed less well-defined binding modes. For ABTS, 33 of the top binding poses were still associated with the protein after 5 ns but the majority (23) were defined by the analysis as unclustered (cluster -1). Reactive blue 4 follows a similar trend.

The results of the BUDE docking calculation show several small molecule binding sites on C45. The stability and promiscuity of substrates bound in the SP pocket within cluster 3, as well as its proximity to the heme, make this an extremely convincing candidate binding site. To validate these computational results, this pocket should be the target for further computational and experimental studies, producing mutants that perturb or enhance the substrate binding in this site. From the frequency of close contact analysis, residues L22, E19 and E47 are the three top candidates for mutation. The fact that this pocket does not appear to be highly selective is in line with very similar experimentally derived K_M values for different substrates (Table 2.3.4).

Cluster 6 is most likely too far from the heme to be kinetically relevant, unless tryptophan radicals on W8 and W80 are formed prior to binding, thus mediating long-range electron transfer. The cluster 5 binding site is within range for rapid, direct electron transfer to the heme and is just 6 Å from the α-heme edge. This could also be a secondary candidate for further computational and experimental study. Clusters 1 and 2 that make direct contact with the heme could also be experimentally relevant, even though they are less frequently sampled computationally. The proximity to the heme edge means that electron transfer would be rapid and even highly transient interactions with electron donating substrates could result in successful electron transfer. This close but transient interaction could also explain the lack of variation in experimentally derived K_M values. As discussed in Section 1.7 most natural peroxidase are generally thought to bind small molecule substrates near the heme edge. These binding results support this trend with a number of ligands remaining in clusters 1 and 2 close to the heme over the duration of the 5 ns MD simulation.

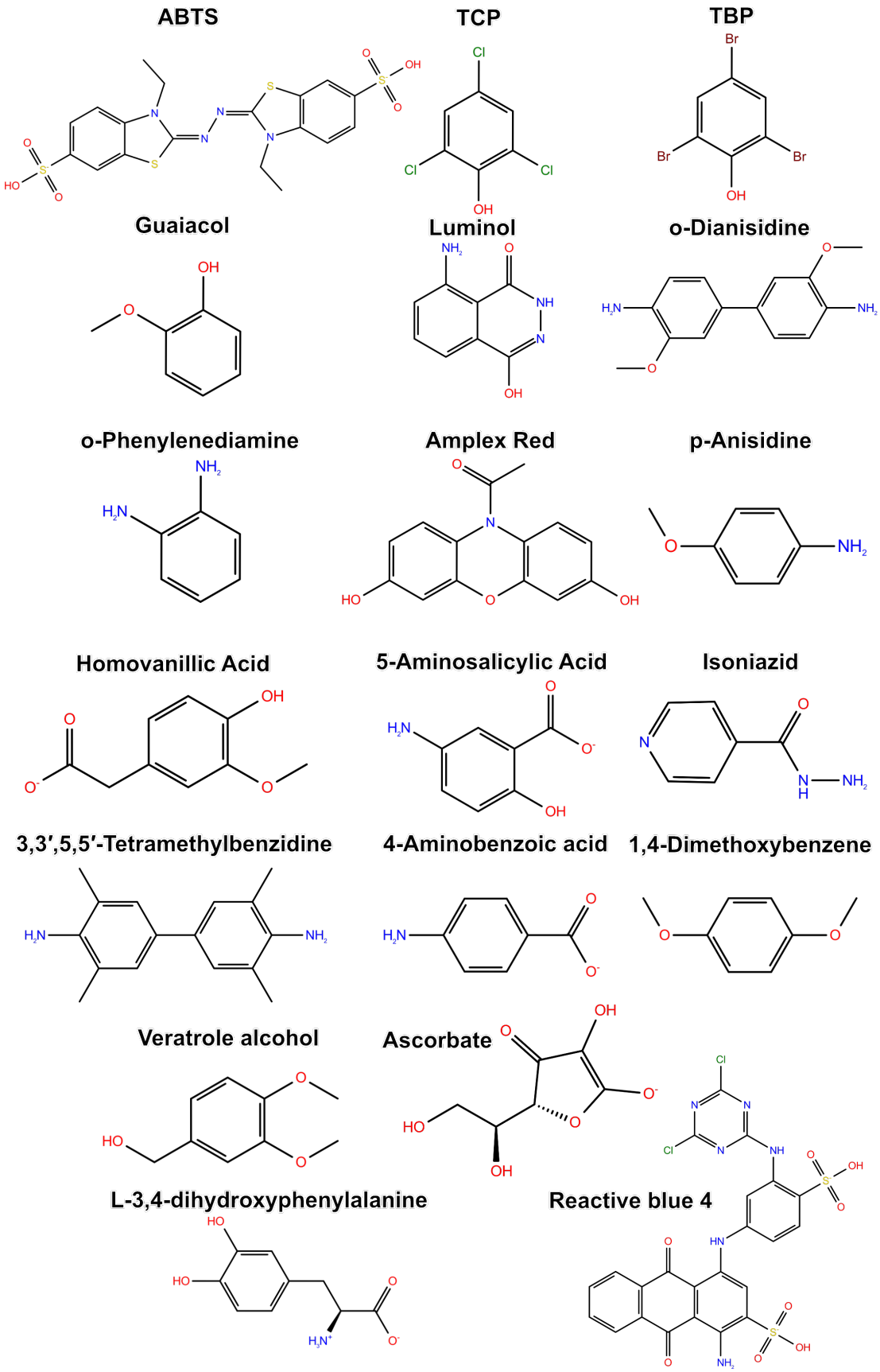


Figure 2.3.17 Structures of small molecule ligands

Structures of ligands used in the computational binding screen protonation states taken at pH 8.6.

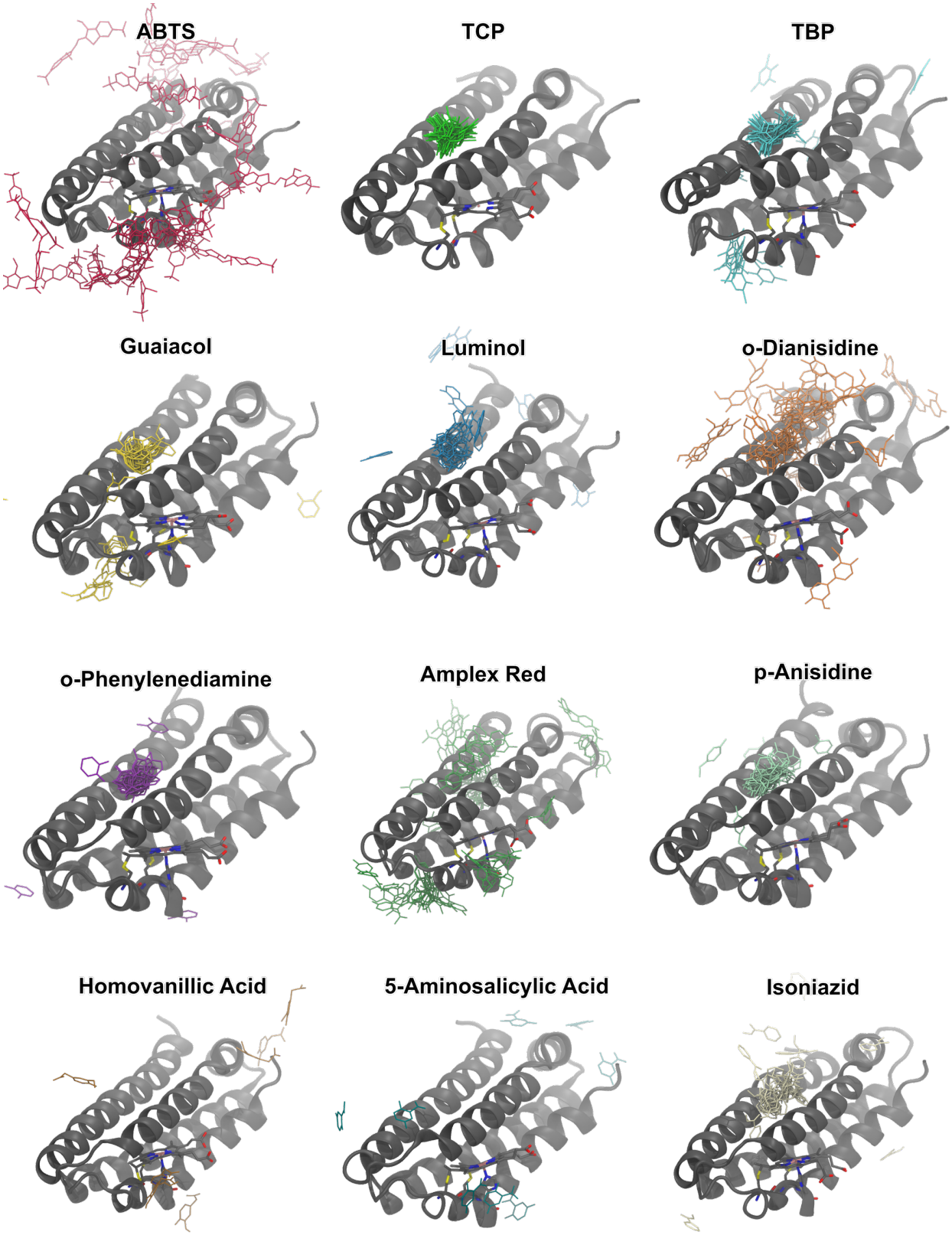


Figure 2.3.18 Binding screen results. *An overlay of ligands associated with protein after 5 ns is shown for each ligand.*

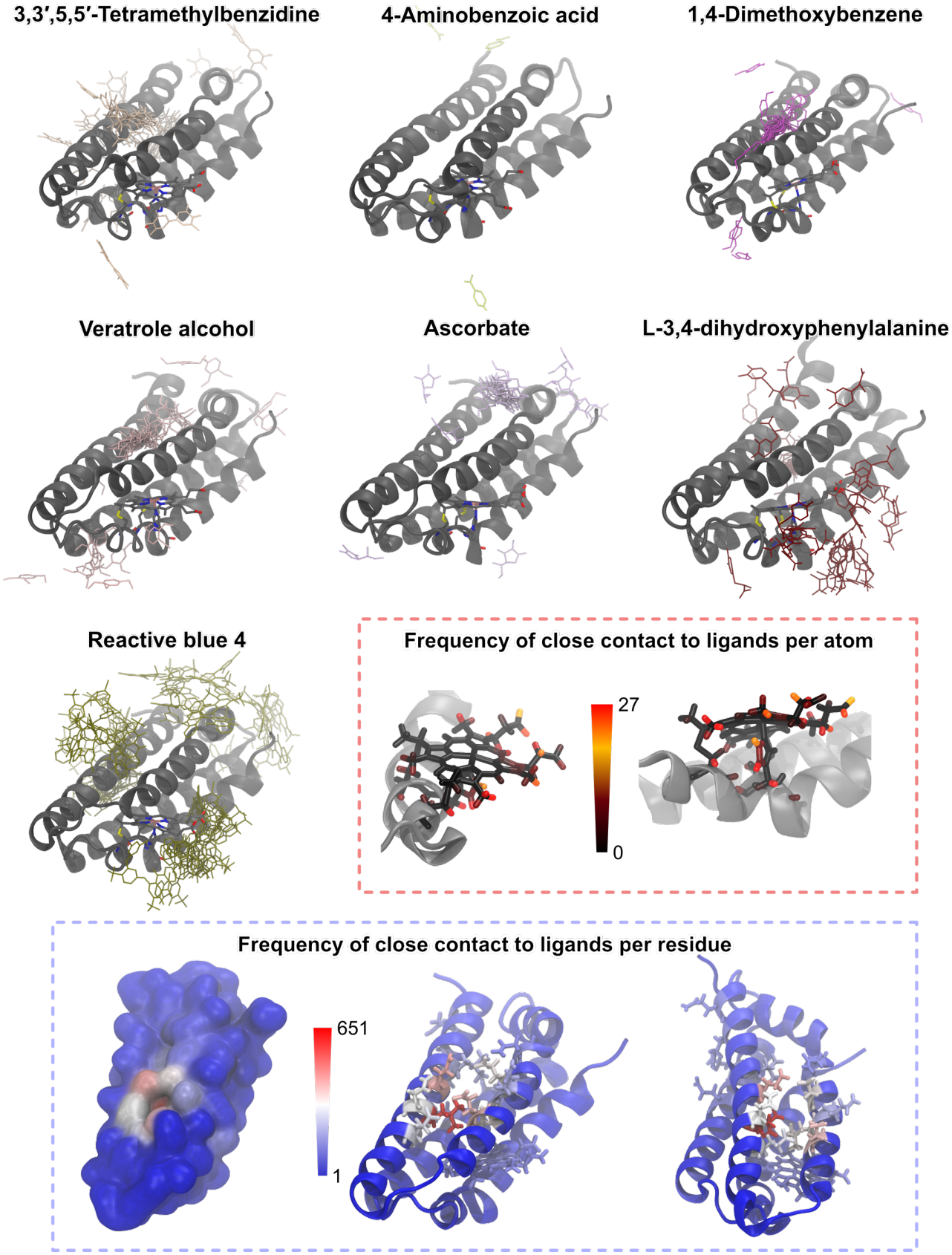


Figure 2.3.19 Binding screen results continued. *An overlay of ligands associated with protein after 5 ns is shown for each ligand. The red and blue boxes show the results of calculating the frequency of close contacts with ligands for the heme atoms and all residues respectively.*

2.3.12 Substrate promiscuity

Like natural peroxidases, C45 oxidises a broad range of organic substrates. This section details the Michaelis-Menten kinetic analysis of a number of substrates that are oxidised by C45 and others for which C45 displays little to no activity. The methods for these experiments are detailed in Section 2.2.12. Unless otherwise stated, reactions with C45, electron-donating substrates and H_2O_2 were carried out in redox buffer (100 mM KCl, 20 mM CHES and pH 8.6), at room temperature (298 K) and with the addition of H_2O_2 (100 μM final concentration) to initiate the reaction. This concentration of H_2O_2 was used to limit heme degradation which is especially evident with poorly reacting substrates and at low substrate concentrations. The observed k_{cat} for these reactions where H_2O_2 is at a rate limiting concentration are denoted as k_{cat}*. Rate limiting concentrations of peroxide meant conducting the experiments to fit to ping-pong reaction mechanism was not possible. High background oxidation of the substrate and product degradation also become problematic at high H_2O_2 concentrations.

The substrates for which it was possible to obtain Michaelis-Menten kinetics for C45/H_2O_2-catalysed oxidation were (in order of redox potential): Cytochrome *c*, *o*-Phenylenediamine, Amplex Red, Guaiacol, Luminol, *o*-Dianisidine, 2,4,6-tribromophenol (TBP) and 2,4,6-trichlorophenol (TCP). Homovanillic acid, *p*-Anisidine and Orange II were oxidised by C45 but experimental complications made it difficult to characterise fully. Homovanillic acid is not stable in redox buffer at pH 8.6. A convenient wavelength for monitoring substrate depletion in *p*-Anisidone and Orange II. 4-Aminobenzoic Acid 1,4-Dimethoxybenzene, Veratrole Alchohol were oxidised poorly or not at all by C45.

2.3.12.1 Cytochrome c

Cytochrome *c* (Cyt *c*) is a small soluble redox-active protein found in the inner mitochondrial membrane of eukaryotes. Its primary role is in the electron transport chain, transferring electrons from the Cytochrome *c*-oxidoreductase (Complex III) to Cytochrome *c* oxidase (Complex IV). At pH 8.3 it has a redox potential of +261 mV (vs SHE) [261]. In *S. cerevisiae* it is the natural substrate of Cytochrome *c* peroxidase (CcP) which oxidises Ferrous-Cyt *c* (Fe^{2+}) to Ferric-Cyt *c* (Fe^{3+}). Over the course of one peroxidase catalytic cycle, two electrons are required from 2 Ferrous-Cyt *c*'s, to reduce both compound I and then II. Figure 2.3.20A shows the general reaction scheme as well as illustrating the complimentary surface

charges of the two proteins. The complementary surfaces may facilitate electron transfer and transient binding.

Reactions carried out in the absence of peroxidase (C45) result in the Cytochrome *c* being slowly oxidised. This background does not significantly impact steady-state experiments at this low concentration of hydrogen peroxide. The steady-state kinetics for ferrous-Cytochrome *c* oxidation are found in Figure 2.3.20.D. The concentration of ferrous-cytochrome was varied between 0 and 50 μM and results fit a simple Michaelis-Menten model. C45 has a $k_{cat(cyt\ c)}$ at limiting H_2O_2 concentration ($k_{cat}*_{(cyt\ c)}$) of 0.51 ± 0.01 s^{-1} and a $K_{M(cyt\ c)}$ of 1.2 ± 0.6 µM resulting in a $k_{cat}*/K_{M(cyt\ c)}$ of 4.3 ± 2.1 x 10^5 $M^{-1}s^{-1}$. This catalytic efficiency (k_{cat}/K_M) is two orders of magnitude lower than the natural enzyme and substrate combination (CcP and Cytochrome *c* pH 6) [262]. A full comparison of enzyme kinetics for all substrates tested is found in Figure 2.3.30.

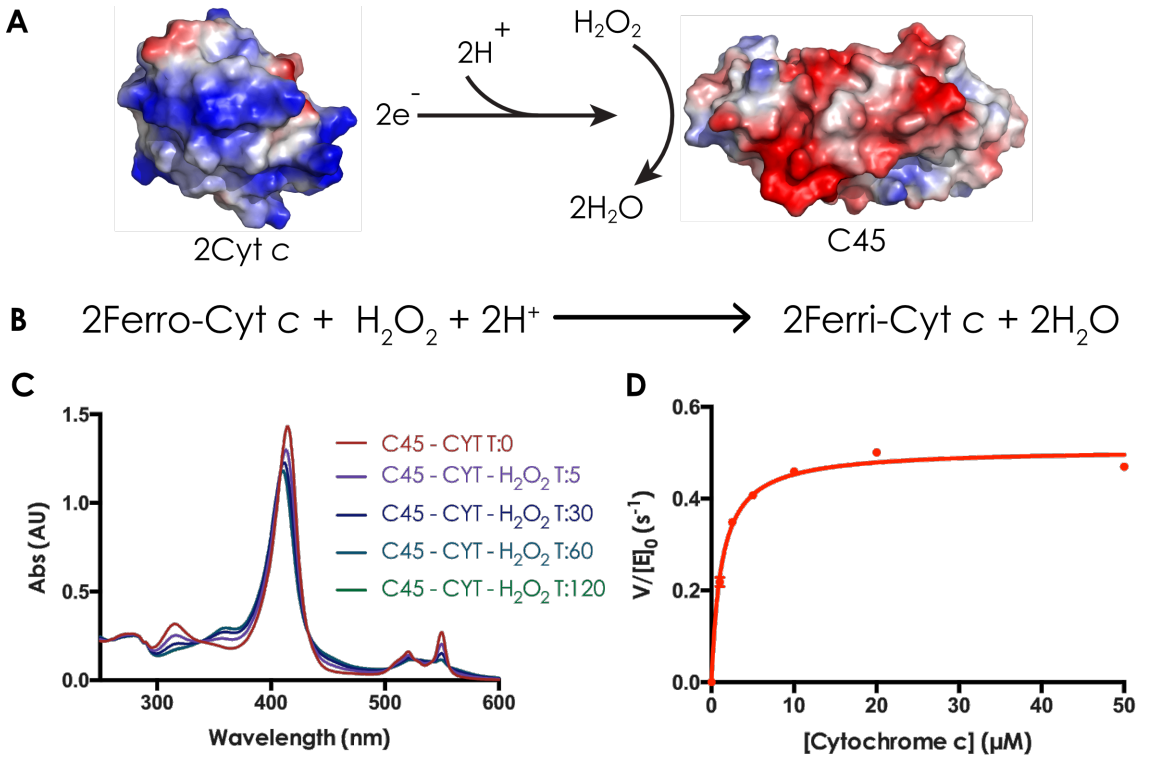


Figure 2.3.20 Cytochrome c oxidation with C45.
(A) *The reaction of Cytochrome c with C45 with protein surface representations showing blue as positive and red as negative.* ***(B)*** *The reaction word equation between ferro-cyctochrome c and C45.* ***(C)*** *UV-visible spectral changes occurring on mixing 250 nM C45, 100 µM Cytochrome c and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.*
(D) *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve. Steady-state data from this figure was collected with Dan Watkins* [70]

2.3.12.2 o-Phenylenediamine

***o*-Phenylenediamine** (oPDA) is a common small molecule peroxidase substrate. Product formation can be monitored spectroscopically at 450 nm where 2,3-diaminophenazine absorbs. Figure 2.3.21.A shows the reaction scheme, two molecules of *o*-Phenylenediamine are required per peroxidase cycle. *o*-Phenylenediamine was stable over the course of the experiment at pH 8.6 and mixing with H_2O_2 in the absence of C45 did not produce any significant oxidation. It has a redox potential of +400 mV at pH 7 [263]. In the computational binding study with C45 (Section 2.3.11) it was predicted to bind close to and above the heme (cluster 3).

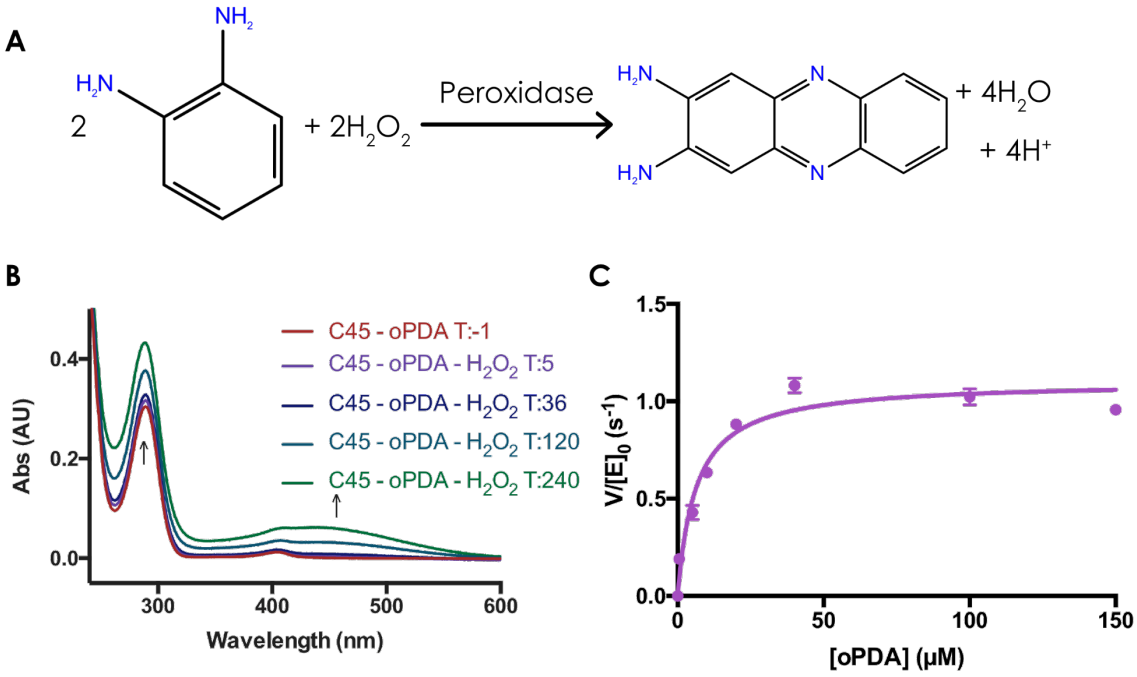


Figure 2.3.21 o-Phenylenediamine oxidation with C45.
(A) *The reaction of 1 o-Phenylenediamine (oPDA) dehaloperoxidase producing primary product 2,3-diaminophenazine* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 100 µM o-Phenylenediamine and 100 µM H_2O_2. "T:x" is the time in seconds from the initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve.*

Steady-state kinetics were collected by monitoring product (2,3-diaminophenazine) formation at 450 nm on a stopped-flow spectrophotometer. The steady-state kinetic analysis of C45 with *o*-phenylenediamine gave a k_{cat}* of 1.1 ± 0.03 (s^{-1}), a K_M of 6.2 ± 0.7 (µM) and k_{cat}*/K_M of 1.8 ± 0.19 x10^5 ($M^{-1}s^{-1}$). The C45's catalytic efficiency of oPDA oxidation is an order of magnitude lower than the reaction with HRP. This is down to C45's low k_{cat}* being matched by a similarly low K_M, whereas HRP has a much higher k_{cat} but also a high K_M.

2.3.12.3 Amplex Red

Amplex Red (AmR) is a peroxidase substrate that, after peroxidase-induced dealkylation, produces a highly fluorescent product resorufin. Resorufin also produces a distinct absorption spectrum allowing the reaction to be followed by the change in absorbance at 572 nm. Amplex Red was stable over the time course of the experiment at pH 8.6 and mixing with H_2O_2 in the absence of C45 did not produce any significant oxidation. It has a redox potential of +600 mV (pH 7.4 [264]). In the computational binding study with C45 (Figure 2.3.16), Amplex red had a high number of binding sites that remained associated with C45 (45/50) after the molecular dynamics simulations. Fourteen of these bound far from the heme (cluster 6), sixteen in clusters close to the heme (clusters 3 and 6), and four associated with the heme edge (cluster 1).

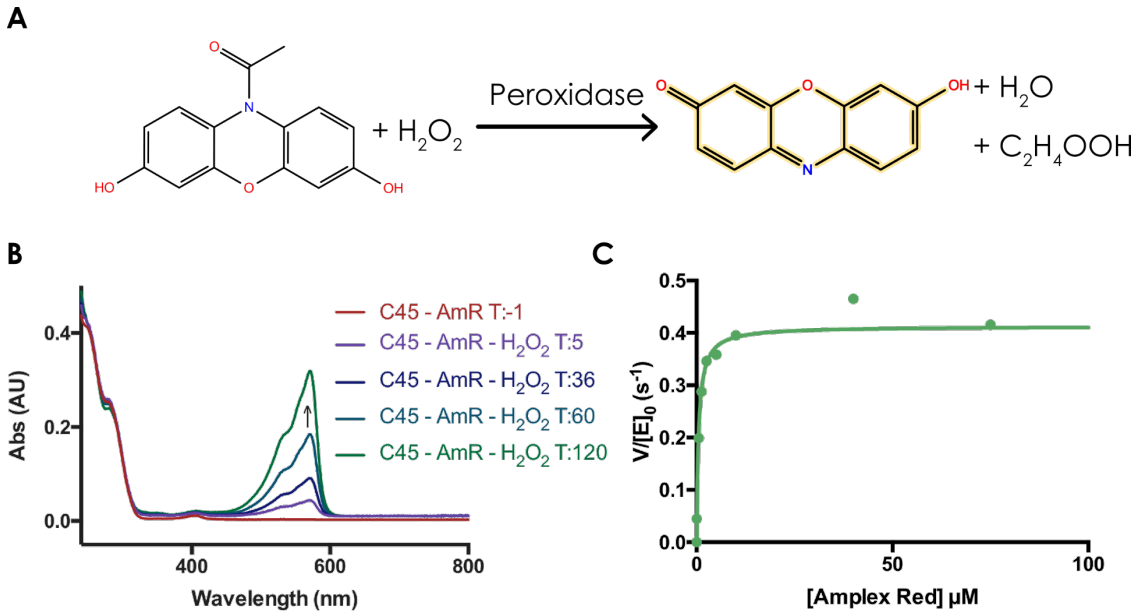


Figure 2.3.22 Amplex Red oxidation with C45

(A) *The reaction of amplex red and C45 producing resorufin* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 75 µM amplex red and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve*

The kinetic analysis of C45 with Amplex Red gave a k_{cat}* of 0.41 ± 0.01 (s^{-1}), a K_M of 0.53 ± 0.04 (µM) and k_{cat}*/K_M of 7.8 ± 0.65 x10^5 ($M^{-1}s^{-1}$). When compared to HRP (k_{cat}*/K_M = 8.1 ± 1.06 x10^6 $M^{-1}s^{-1}$, pH 7.4) the catalytic efficiency is an order of magnitude lower [265].

2.3.12.4 Guaiacol

Guaiacol is a small phenolic molecule that has a rich history as a peroxidase substrate. The production of tetraguaiacol is monitored by the absorbance change at 470 nm. Guaiacol was stable over the time course of the experiment at pH 8.6 and mixing with H_2O_2 in the absence of C45 did not produce any significant oxidation. It has a redox potential of +680 mV (pH 7 [266]). In the computational binding study with C45 (Figure 2.3.16), guaiacol had 36/50 binding sites that remained associated with C45 after MD. The majority of these (23), bound close to and above the heme in cluster 3.

The kinetic analysis of C45 with Guaiacol gave a k_{cat}* of 0.76 ± 0.01 (s^{-1}), a K_M of 2.6 ± 0.3 (µM) and k_{cat}*/K_M of 2.9 ± 0.33 x10^5 ($M^{-1}s^{-1}$). Having a very low K_M means that it outperforms CcP and AXP k_{cat}*/K_M of (7.7 ± 1.04 x10^4, 3.2 ± 0.43 x10^3 $M^{-1}s^{-1}$) in terms of catalytic efficiency by at least an order of magnitude. Despite a k_{cat}* three orders of magnitude lower, C45 matches the catalytic efficiency of HRP (k_{cat}*/K_M 1.4 ± 0.37 x10^5).

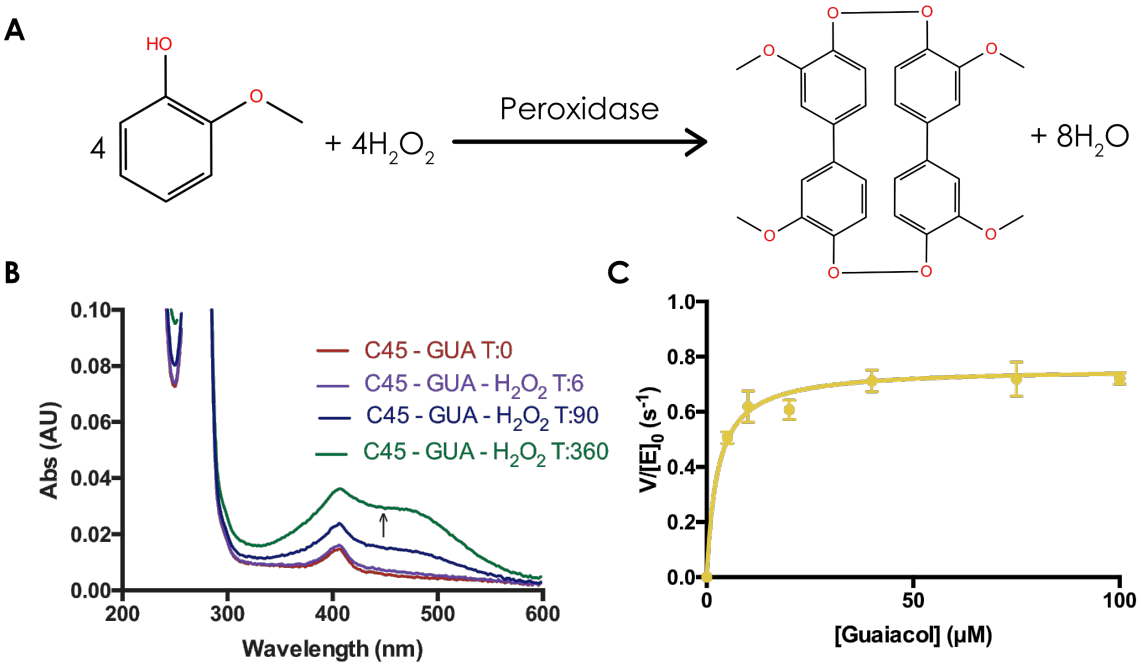


Figure 2.3.23 Guaiacol oxidation with C45.
(A) *The reaction of guaiacol peroxidases producing primary product tetraguaiacol* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 100 µM guaiacol and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve.*

2.3.12.5 Luminol

Luminol is used in crime scene forensics, as it is oxidised *via* a chemiluminescent process when it comes into contact with blood and other bodily fluids [267]. It was stable in redox reaction buffer over the course of the experiment and shows minimal oxidation when mixed with H_2O_2 without C45. To quantify the concentration of product using the chemiluminescent signal requires extensive calibration of the detector. As a result, the absorbance changes associated with substrate loss were used to follow the reaction. Luminol has a redox potential of +750 mV (pH 9 [268]). The computational binding study with C45 (Figure 2.3.16) showed luminol had 22 binding sites that remained associated with C45 after MD. Five were un-clustered (cluster -1) but over 15 Å from the heme iron, 17 were close to and above the heme in cluster 3.

The kinetic analysis of C45 with luminol gave a k_{cat}* of 0.65 ± 0.03 (s^{-1}), a K_M of 2.9 ± 0.7 (µM) and k_{cat}*/K_M of 2.3 ± 0.52 x10^5 ($M^{-1}s^{-1}$). Once again the observation of a low K_M for C45 when fitting to Michaelis–Menten kinetics means that C45 and HRP are in the same order of magnitude when it comes to catalytic efficiency (HRP k_{cat}*/K_M = 1.9 ± 0.14 x10^5 $M^{-1}s^{-1}$ [269]).

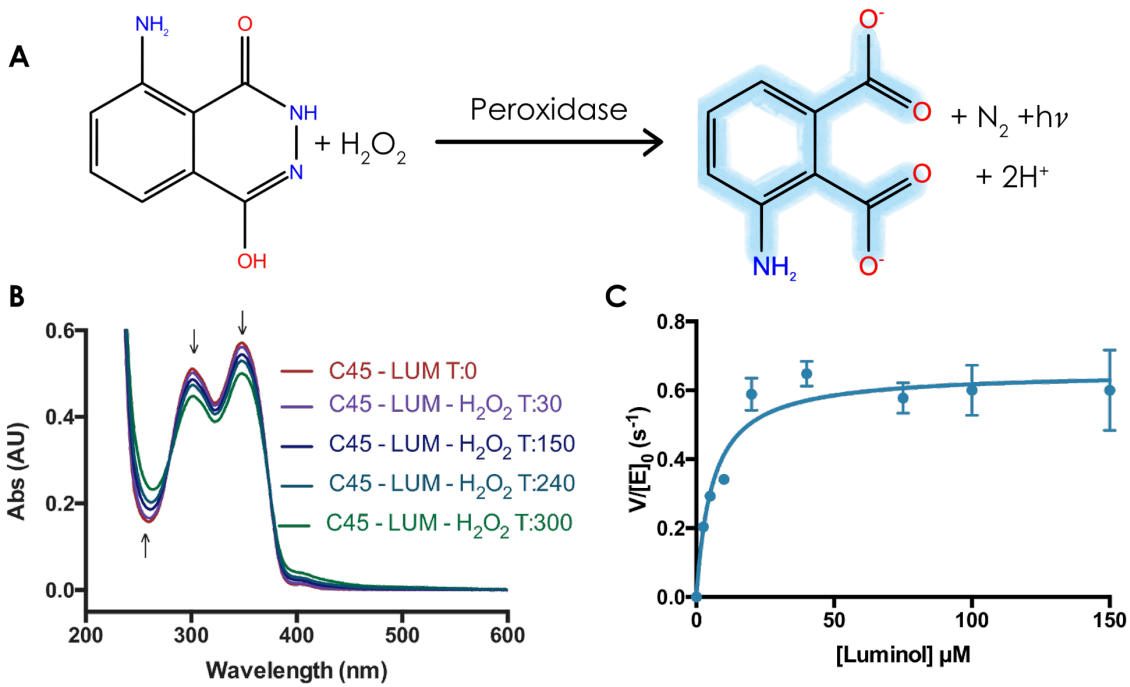


Figure 2.3.24 Luminol oxidation with C45

(A) *The reaction of luminol peroxidases.* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 75 µM luminol and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve.*

2.3.12.6 o-Dianisidine

***o*-Dianisidine** (oDi) can be used as an azo dye-precursor as well as being used as a substrate in colorimetric peroxidase assays. It was stable in redox reaction buffer (pH 8.6) over the course of the experiment and shows minimal oxidation when mixed with H_2O_2 without C45. The absorbance changes associated with the formation of the oxidised product were used to follow the reaction. *o*-Dianisidine has a redox potential of +809 mV (in water [270]). Figure 2.3.16 from the computational binding study showed *o*-Dianisidine had 38 binding sites that remained associated with C45 after MD. 19 found in cluster 3 in and around the same small phenolic binding pocket above the heme. 4 were clustered group 6 on the other end of the maquette far from the heme. 2 of the 9 un-clustered (cluster -1) were within 15 Å of the heme iron. The kinetic analysis of C45 with *o*-Dianisidine gave a k_{cat}* of 1 ± 0.08 (s^{-1}), a K_M of 1.9 ± 0.42 (µM) and k_{cat}*/K_M of 5.4 ± 1.26 x10^5 ($M^{-1}s^{-1}$).

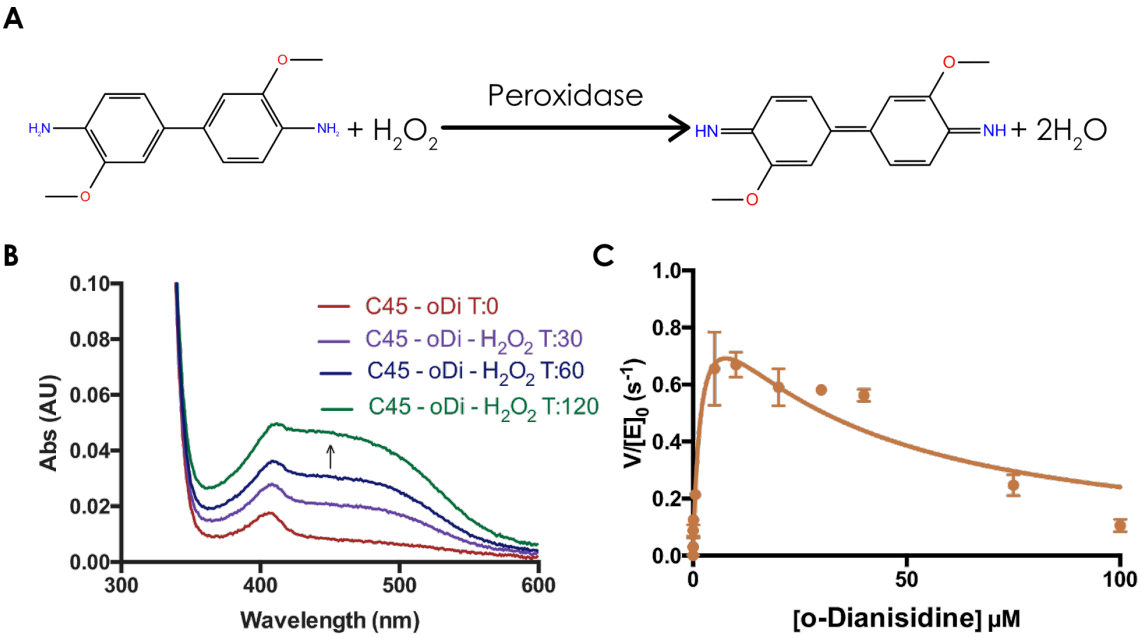


Figure 2.3.25 o-Dianisidine oxidation with C45

(A) *The oxidation of o-Dianisidine by peroxidases.* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 40 µM o-Dianisidine and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 µM H_2O_2, fitted to a Michaelis–Menten curve.*

2.3.12.7 TCP and TBP

2,4,6-trichlorophenol **(TCP)** and 2,4,6-tribromophenol **(TBP)** are halogenated phenol derivatives whose environmental impact is of growing concern [189]. They occur as intermediates in industrial synthesis, in the breakdown of flame retarded products and some organisms release them as secondary metabolites to eliminate competition in their environmental niche [188]. A worm-like terebellid polychaete (*Amphitrite ornate*) cohabits among organisms that secrete bromophenols, so expresses a dehaloperoxidase (DHP) as a response [180]. See Section 1.7.4 for details on DHP and its two isoforms (DHP a and DHP B). The results of computation binding study show all of the highest energy TCP binding positions are in cluster 3, the pocket close to and above the heme. For TBP the majority of the binding poses that remained after MD were in the same position as TCP with only 11 associating at other positions around the protein (Figure 2.3.16).

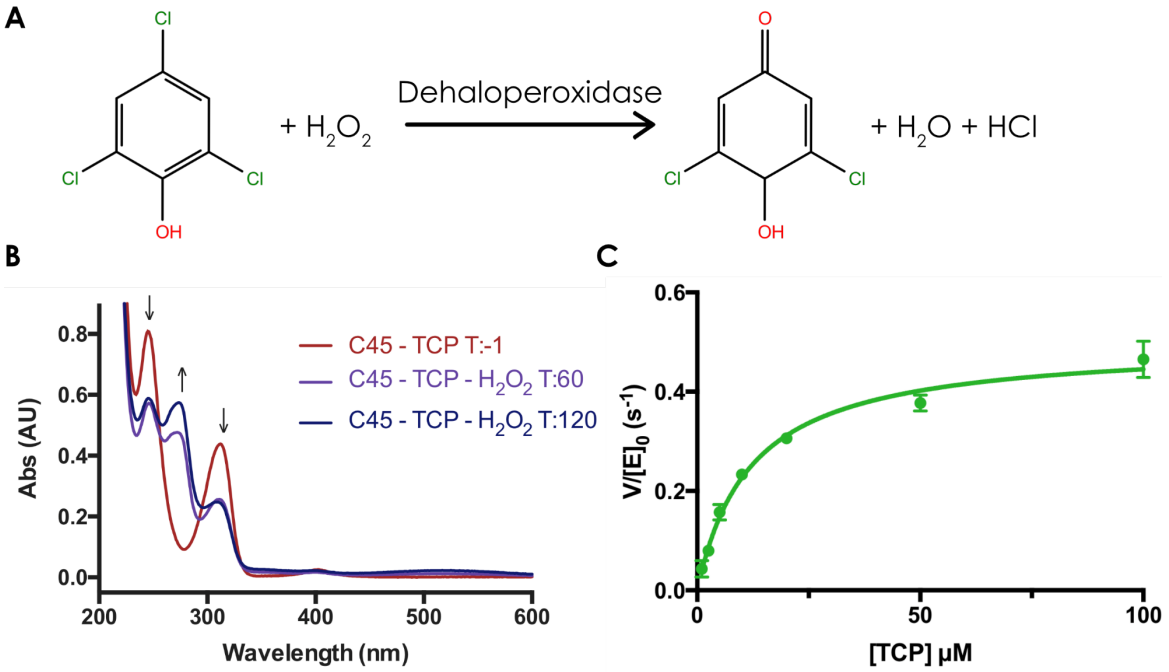


Figure 2.3.26 TCP oxidation with C45

(A) *The reaction of TCP dehaloperoxidase producing primary product 2,6-dichloroquinone.* ***(B)*** *UV-visible spectral changes occurring on mixing 250 nM C45, 100 µM TCP and 100 µM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at 100 µM H2O2 (rate-limiting), fitted to Michaelis–Menten curve*

The absorbance changes associated with TCP and TBP oxidation to their primary products 2,6-dichloro-1,4-benzoquinone (DCQ) and 2,6-dibromo-1,4-benzoquinone (DBQ)

are seen in panel A of Figure 2.3.26 and Figure 2.3.27. Spectra of both were stable over the duration of the experiments at pH 8.6. There was no evidence of oxidation of either when mixing with H_2O_2 in the absence of C45.

The kinetic analysis of C45 with TCP gave a k_{cat}* of 0.5 ± 0.01 (s^{-1}), a K_M of 12 ± 1.1 (μM) and k_{cat}*/K_M of 4.2 ± 0.40 x10^4 ($M^{-1}s^{-1}$). For TBP C45 has a k_{cat} * of 0.43 ± 0.01 (s^{-1}), a KM of 11 ± 0.70 (μM) and k_{cat}*/K_M of 4 ± 0.25 x10^4 ($M^{-1}s^{-1}$). Table 2.3.4 shows that in terms of catalytic efficiency C45 beats the natural DHP A and matches DHP B isoform. The results with C45 were at limiting H_2O_2 concentration (100 μM) when the kinetics for DHP were taken at 160 μM H_2O_2.

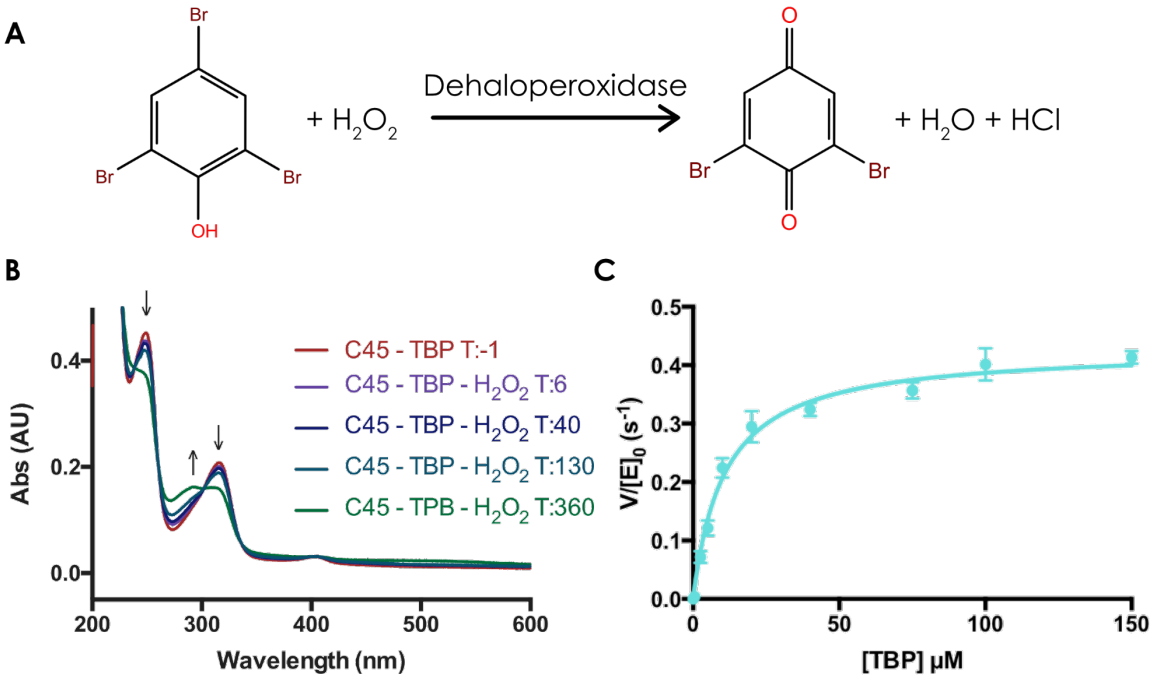


Figure 2.3.27 TBP oxidation with C45

(A) *The reaction of TBP dehaloperoxidase producing primary product 2,6-dibromoquinone* ***(B)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 40 μM TBP and 100 μM H_2O_2. "T:x" is the time in seconds from initial mixing.* ***(C)*** *Steady-state kinetics at (rate-limiting) 100 μM H_2O_2, fitted to a Michaelis–Menten curve*

2.3.12.8 Other substrates turned over by C45

Homovanillic acid (HVA), *p*-Anisidine (pAN) and Orange II (ORA) are documented peroxidase substrates with redox potentials of +430, +766 and +868 mV respectively [271–273]. While they were all oxidised by C45, a full kinetic characterisation was not carried out owing to practical complications. Homovanillic acid is sensitive to light and was unstable at pH 8.6. However, it was clear to see that with mixing with C45 and H_2O_2 turnover was increased relative to the background rate of decomposition by at least 2 orders of magnitude. There was no bleaching of the heme Soret λ_{max}, which occurs when mixing C45 and H_2O_2 in the absence of an electron donating substrate, indicating that homovanillic acid can be oxidised by C45. *p*-Anisidine and Orange II were stable at pH 8 but it was not possible to obtain UV-visible extinction coefficients for their oxidation products; and measuring substrate decrease proved challenging at higher concentrations. Figure 2.3.28 shows the UV-visible changes after oxidation with C45 as well as the structure of the substrates.

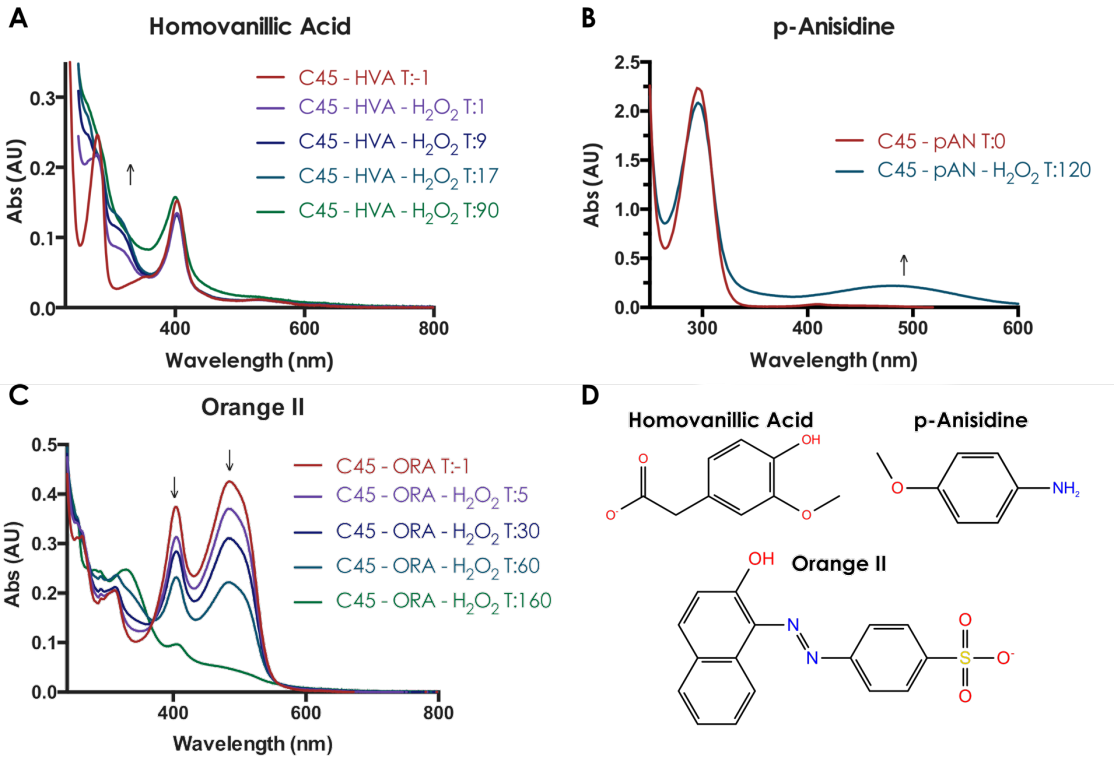


Figure 2.3.28 Substrates without steady-state kinetics.
(A) *UV-visible spectral changes on mixing 700 nM C45, 100 μM Homocanillic acid and 100 μM H_2O_2.* ***(B)*** *UV-visible spectral changes occurring on mixing 250 nM C45, 100 μM p-Anisidine and 100 μM H_2O_2.* ***(C)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 40 μM Orange II and 100 μM H_2O_2.* ***(D)*** *Skeletal structures of small molecule substrates. "T:x" is the time in seconds from initial mixing.*

2.3.12.9 Substrates not oxidised by C45

4-Aminobenzoic acid (4ABA), 1,4-dimethoxybenzene (DEOB) and veratrole alcohol (VA) all have relatively a high oxidation potentials and +1080, +1300 and +1360 mV [272, 274, 275]. All are oxidised by peroxidases with high redox potentials, like Lignin Peroxidase (LiP). Veratrole alcohol is produced as a secondary metabolite by white-rot basidiomycetous fungi and is thought to act as an intermediate in the degradation of lignin (see in Section 1.7.2) [276]. Figure 2.3.29 shows the UV-visible spectra of mixing these compounds with C45 and H_2O_2. In all cases substrates were stable with C45 at pH8.6 and reactions were initiated be 5 μl of H_2O_2. The T:-1 trace in panel A,B and C is before the addition of H_2O_2. For all substrates there are clear spectral changes corresponding to heme degradation of C45 (decrease at 406 nm). This is the same phenomenon observed when mixing C45 and H_2O_2 in the absence of an electron donating substrate, the only exception being the very slight reduction in the absorbance peak of 4-Aminobenzoic acid.

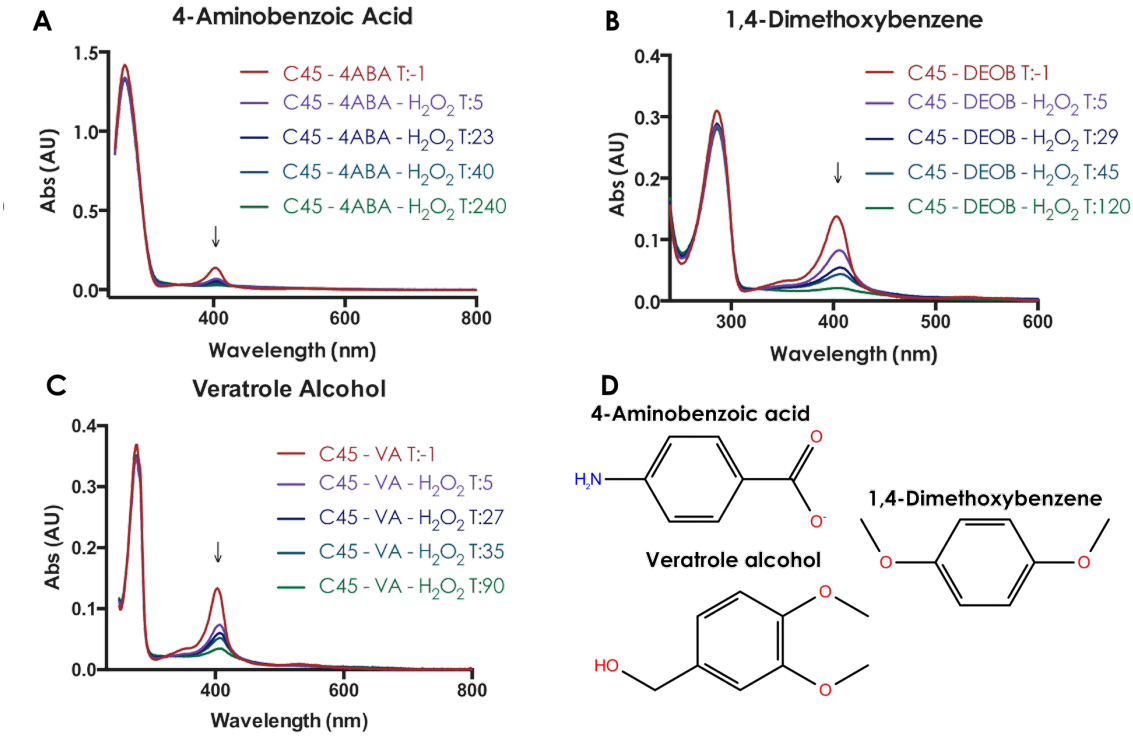


Figure 2.3.29 Substrates with very poor or no turnover.

(A) *UV-visible spectral changes on mixing 100 nM C45, 100 μM 4-Aminobenzoic acid and 100 μM H_2O_2.* ***(B)*** *UV-visible spectral changes occurring on mixing 250 nM C45, 100 μM 1,4-Dimethoxybenzene and 100 μM H_2O_2.* ***(C)*** *UV-visible spectral changes occurring on mixing 100 nM C45, 40 μM Veratrole Alcohol and 100 μM H_2O_2.* ***(D)*** *Skeletal structures of small molecule substrates. "T:x" is the time in seconds from initial mixing. In all traces the reduction at 405 nm is a result of heme degradation.*

4-Aminobenzoic acid performed relatively poorly in the computational binding study with only 3/50 top binding poses remaining associated with the protein after MD. 1,4-dimethoxybenzene and VA exhibited very similar results, with the majority of top biding poses remaining associated with the protein after MD (41/50 and 45/50). Figure 2.3.16 shows that the majority of these, in both cases, were found in the cluster 3 close to and above the heme.

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	Condition
C45	Cyt c	0.51* ± 0.01	1.2 ± 0.6	4.3 ± 2.13 x10^5	pH 8.6
CcP	Cyt c	1510 ± 259	93 ± 31	1.6 ± 0.61 x10^7	pH 6 [262]
C45	oPDA	1.1* ± 0.03	6.2 ± 0.7	1.8 ± 0.19 x10^5	pH 8.6
HRP	oPDA	1860 ± 43	600 ± 200	3.1 ± 1.03 x10^6	pH 6 [277]
C45	Amplex Red	0.41* ± 0.01	0.53 ± 0.04	7.8 ± 0.65 x10^5	pH 8.6
HRP	Amplex Red	860 ± 38	106 ± 13	8.1 ± 1.06 x10^6	pH 7.4 [265]
C45	Guaiacol	0.76* ±0.01	2.6 ± 0.3	2.9 ± 0.33 x10^5	pH 8.6
CcP	Guaiacol	4.1 ± 0.3	53 ± 6	7.7 ± 1.04 x10^4	pH 6 [262]
AXP	Guaiacol	41 ± 5	13000 ± 800	3.2 ± 0.43 x10^3	pH 7.4 [278]
HRP	Guaiacol	370 ± 30	2700 ± 700	1.4 ± 0.37 x10^5	pH 7.4 [278]
C45	Luminol	0.65* ± 0.03	2.9 ± 0.7	2.3 ± 0.52 x10^5	pH 8.6
HRP	Luminol	275 ±10	1425 ± 93	19 ± 0.14 x10^5	pH 5 [269]
C45	H_2O_2 (ABTS)	1200 ± 20	94 ± 6	1.3 ± 0.51 X10^4	pH 8.6
C45	ABTS	1200 ± 20	379 ± 25	3.2 ± 0.43 x10^6	pH 8.6
HRPc	ABTS	4100	800	5.1 x10^6	pH 4.6 [182]
HRP	ABTS	332 ± 18	233 ± 21	1.4 ± 0.15 x10^6	pH 5 [265]
C45	o-Dianisidine	1.0* ± 0.08	1.9 ± 0.42	5.4 ± 1.26 x10^5	pH 8.6
C45	TCP	0.50* ± 0.01	12 ± 1.1	4.2 ± 0.40 x10^4	pH 8.6
DHP A	TCP	8.1 ± 0.3	1004 ± 90	8.1 ± 0.78 x10^3	pH 7 [181]
DHP B	TCP	14.1 ± 0.7	297 ± 35	4.7 ± 0.61 x10^4	pH 7 [181]
C45	TBP	0.43* ± 0.01	11 ± 0.70	4.0 ± 0.25 x10^4	pH 8.6
DHP A	TBP	1.81 ± 0.14	641 ± 84	2.8 ± 0.43 x10^3	pH 7 [181]
DHP B	TBP	2.04 ± 0.1	91.9 ± 16.2	2.2 ± 0.41 x10^4	pH 7 [181]

Table 2.3.4 Comparison of kinetics results.

*Showing Michaelis–Menten kinetic fit results for each electron donating substrate (column 2) with various peroxidases (column 1). The results fro C45 are shaded in blue. * result is k_{cat}* obtained at limiting H_2O_2 with 100 nM C45 and 100 μM H_2O_2. Values for other enzymes are not shaded and have their literature reference in the last column. The ± errors quoted are standard deviation.*

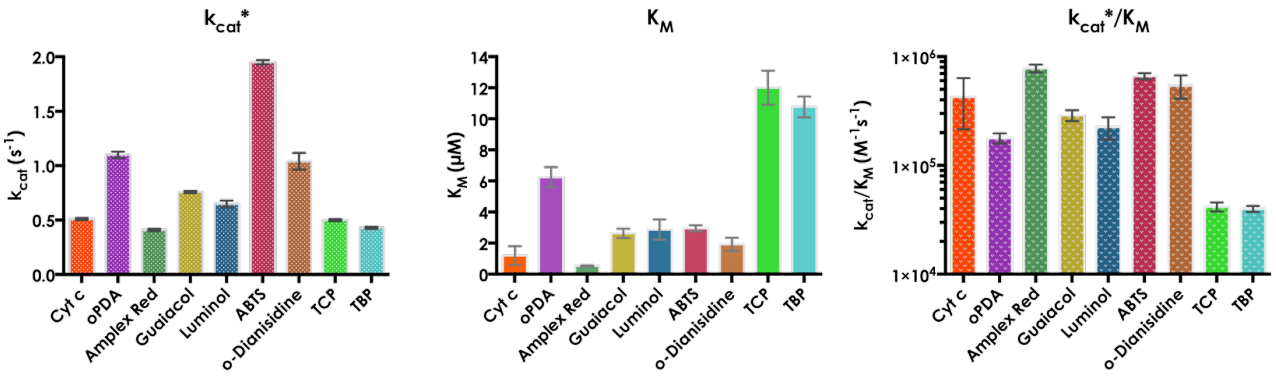


Figure 2.3.30 C45 of kinetics results.

***(A)** Graph of C45 k_{cat}* values at rate limiting H_2O_2. **(B)** Graph of C45 K_M values for various substrates **(C)** Graph of C45 $k_{cat}*/K_M$ values for various substrates*

	E^0 oxidation potential (mV vs SHE)	Conditions	REF
CcP Fe^{3+}/Fe^{2+}	-234	pH 7	[100]
P450cam Fe^{3+}/Fe^{2+}	-216	pH 7.4	[99]
CYP119 Fe^{3+}/Fe^{2+}	-214	pH 7	[99]
APX Fe^{3+}/Fe^{2+}	-206	pH 7	[101]
C45 Fe^{3+}/Fe^{2+}	-174	pH 8.6	[70]
HRP Fe^{3+}/Fe^{2+}	-136	pH 7.0	[98]
Lignin peroxidase Fe^{3+}/Fe^{2+}	-126	pH 7.0	[103]
MnP Fe^{3+}/Fe^{2+}	-122	pH 7.0	[103]
DHP B Fe^{3+}/Fe^{2+}	206	pH 7	[105]
Cytochrome c Fe^{3+}/Fe^{2+}	261	pH 8.3	[261]
o-Phenylenediamine	400	pH 7.0	[263]
Homovanilic Acid	430	pH 7.0	[271]
Amplex Red	600	pH 7.4	[264]
Guaiacol	680	pH 7	[266]
Isoniazid	~700	Indirect [a]	[279]
Luminol	750	pH 9	[268]
ABTS	760	pH 4	[280]
p-Anisidine	766	**pH 5.1**	[272]
o-Dianisidine	809	In water	[270]
Orange 2	868	pH 8	[273]
TCP	930	50% IPA **pH 5.1**	[281]
HRP CPDI/CPDII	949	pH 7.0	[282]
TBP	984	pH 6 90% Ethanol	[283]
HRP CPDII/Fe^{3+}	991	pH 7.0	[282]
C45 CPDI/CPDII	~1000	Indirect pH 8.6 [b]	
4-Aminobenzoic Acid	1080	pH 5.1	[272]
CcP CPDI/CPDII	1087	pH5.3	[108]
HRP CPDI/CPDII	1120	indirect low pH	[169]
APX CPDI/CPDII	1156	pH 7	[101]
1,4-Dimethoxybenzene	1300	pH 4.5	[274]
MnP CPDI/CPDII	1300	pH <4	[98]
Veratrole Alchohol	1360	pH 3.6	[275]
LiP CPDI/CPDII	~1450	Indirect [c]	[169]
P450's CPDI/CPDII	1500-2000	[d]	[109]

Table 2.3.5 Oxidation Potentials

Table gives oxidation potential of various small molecules and enzyme species vs the standard hydrogen electrode (SHE). Blue shading denotes potential peroxidase substrate single electron oxidation potentials. Potentials with compound I or II have been abbreviated to CPD I and CPD II. [b] C45 CPDI/CPDII is an estimate based on the substrates it will oxidise. The last column is the literature reference (REF). [a] Obtained though mediator. [c] Inferred though reactivity with a series of substrates. [d] P450 CPDI oxidation potentials are difficult to measure directly but inferred through reactivity.

2.3.13 Live Cell Assays

These experiments aimed to test the reactivity of C45 *in vivo*. There appeared to be no indication of ABTS, luminol or Guaiacol oxidation in live cells when compared to a control of cells expressing a variant of C45 with no heme, C45-NH. Only *o*-phenylenediamine displayed any significant spectroscopic changes showing product formation (2,3-diaminophenazine) after 90 min. *o*-Phenylenediamine oxidation with C45 expressing cells was significantly greater than the control. Figure 2.3.31-E and F show the reaction of the samples with ABTS after lysis of the cells. C45 expressing cell lysate will readily oxidise ABTS whereas the cells expressing C45-NH cannot.

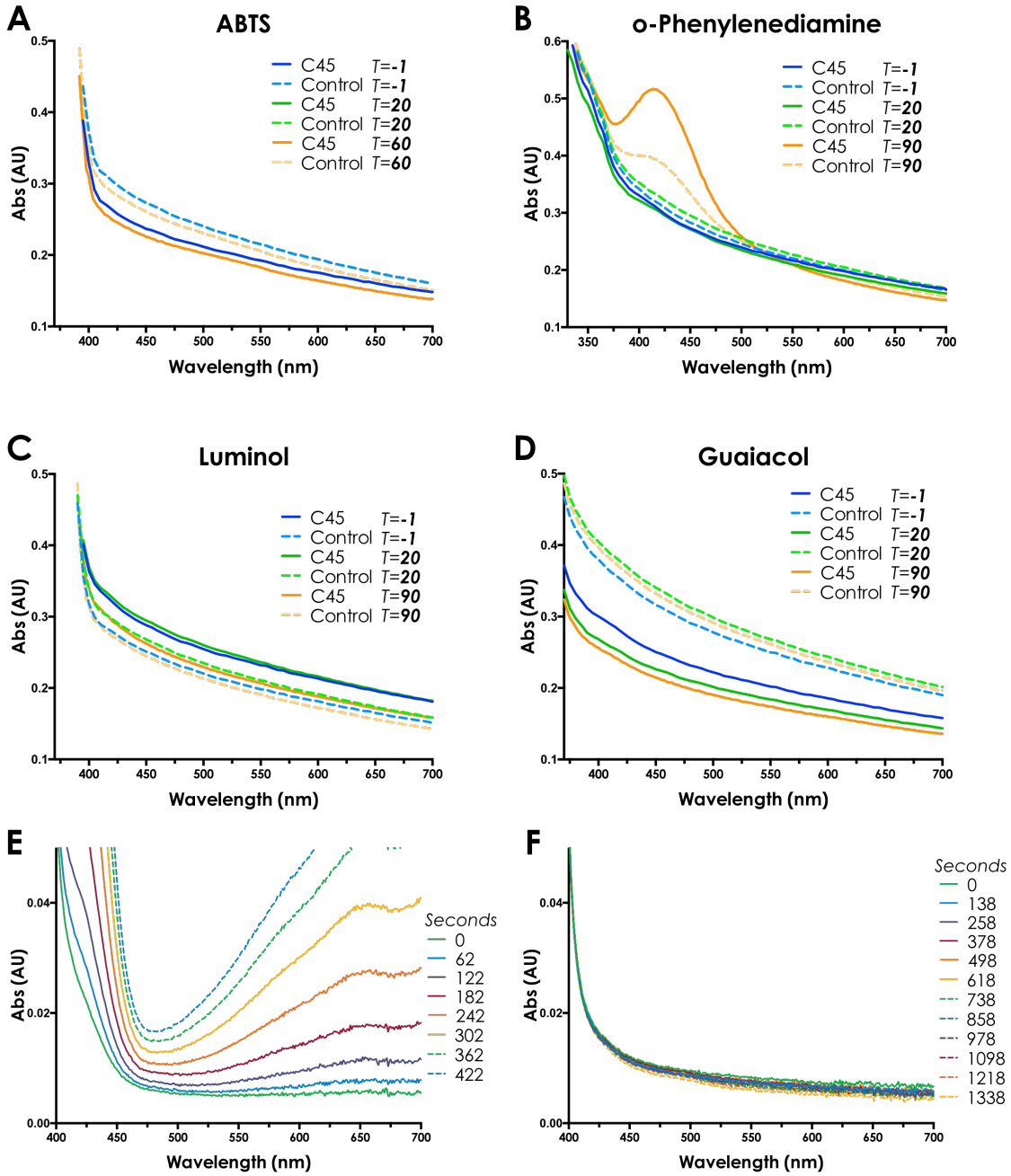


Figure 2.3.31 Live cell Assays.

*Panels **A** to **D** show the reactions with ABTS, o-phenylenediamine, Luminol and Guaiacol at 1 mM and H_2O_2 at 2 mM with 30 ul of live cells. The lines show an average of 4 repeated experiments. The control is cells expressing C45-NH (with no heme binding site) shown with dashed lines. The time of each scan is denoted as T=x in seconds. **E** shows positive control with ABTS, H_2O_2 and lysed cells that were expressing C45. **F** the spetra on mixing with ABTS, H_2O_2 and lysed cells that were expressing C45-NH.*

2.4 Discussion

C45 represents a step-change in the maquette protein design methodology, as it is the first to achieve a level of catalytic activity that places it in an exclusive club of fully *de novo* enzymes. While questions remain about the exact nature of its structure and mechanism, these do not diminish from the broader triumph for rational design that it signifies. In contrast to most other protein design methodologies, which are generally structurally driven, the design of C45 was unashamedly activity driven. Incorporating a *c*-type cytochrome heme cofactor in the *de novo* four-helix bundle allows the protein to be expressed as a holoenzyme *in vivo*. Taking inspiration from natural peroxidases that have a penta-coordinate heme ligation, the last design step for C45 was to remove the distal ligating histidine, leaving the final iron coordination site accessible to water. This modification transformed an unreactive parent protein into a highly catalytically efficient thermostable *de novo* enzyme.

2.4.1 Structure

C45 has a highly repetitive sequence and a significant degree of conformational flexibility. This has limited the structural information that can be gained from NMR and renders attempts at crystallographic studies futile. The NMR results in combination with CD spectroscopy show C45 to be similar to other reported maquette proteins [284]. C45 has clearly has defined secondary structure, randomly coiled loops and some amount of core packing. MD simulations show the model of C45 to have a stable core with highly dynamic loops and surface sidechains.

The results of the Rosetta *ab initio* fold calculations do not show convergence on a single structure. The simulation using constrains to bind heme produced many physically unrealistic high-energy results. The folding simulations without heme more reliably produced four-helix bundles but are purely speculative as it is unknown if they would be able to accommodate the heme. The general result that these preliminary calculations highlight is that by removing the distal heme ligation it now may be possible for two different topologies to support realistic heme orientations. Figure 2.3.9.1 and 5 show those two topologies. Figure 2.3.9.1 is the same overall structure as the existing C45 model which is based on the previous generations of maquettes. In Figure 2.3.9.5 helix 2 and 3, which are linked diagonally across the barrel by a longer loop, have swapped positions. The first step in investigating these topologies would be to run molecular dynamics simulations to see if they are stable and

realistic. The structures generated by Rosetta *ab initio* folding with heme contain a large number of unfavorable backbone torsions so would need to be very rigorously minimized. If this was not possible the helical arrangements could be used as starting points for Rosetta flexible backbone packing software to generate better starting structures for MD. Florescence quenching experiments with mutants of C45 could be designed to test which of the topologies predominates experimentally [285].

To be structurally complete, C45 must incorporate heme. For C45, the covalent attachment of heme is mediated by Cytochrome *c* biogenesis system I (Ccm) in *E. coli*. Work leading up to the production of C45 has shown that the Ccm is able to accommodate non-natural protein substrates [111]. The exact limits of what types of non-natural substrates the Ccm will process have not been studied in depth. However, work with *c*-type binding maquettes has suggested that flexibility of the maquette scaffold may be important for efficient recognition of the CXXCH heme binding motif [69].

Flexibility of the maquette scaffold is likely to be an important factor in the peroxide binding. In majority of natural peroxidases, the distal heme site is accessible to solvent and peroxide but not the small molecule electron donor substrates [152]. The MD simulations of C45 do not sample any conformations where water is bound above the heme, but it is experimentally clear that this is probably the case. One MD result did hint at this possibility (Figure 2.3.8.C). Further investigations of simulations that force water or peroxide binding at the heme are needed to investigate this further. MD simulations of C45 in the compound I and II states could also help to better rationalize what structural changes occur in C45 to accommodate these highly-charged species. The current hypothesis is that flexibility in the second and forth helix allows for the peroxide to bind and be activated at the heme.

2.4.2 What can be learned from pH dependent activity

The UV-visible spectra of C45 in the ferrous form (Figure 2.3.3) show a split Soret λ_{max}, indicating more than one heme iron environment or electronic state. This split peak profile is similar to monohistidine-ligated Cytochrome *c'* and distal mutants of Cytochrome *c* [286, 287]. Split Soret maximum peaks can be explained by either a quantum admixture of high and low-spin states at the heme iron, or mixed populations of protonation states in the titratable water ligand occupying the distal site (H_2O to OH^-), or a combination of both [288, 289]. Figure 2.3.13 shows that C45 transitions between two Soret maxima, in the neutral to

alkaline range, with a pK_a of 8.1. This pK_a value of 'the acid–alkaline transition' is sometimes referred to as pK_a^{AA}. Horse heart myoglobin and DHP display very similar transitions (pK_a^{AA} 8.9 and 8.1) caused by changes in the protonation state of the water bound to the heme [180, 290]. Changes in pH also have a profound effect on the redox potential of C45, moving from -208 mV at pH 10, to -174 at pH 8.6. Two redox potentials are observed at pH 7.5 (-187 and – 126 mV) which are likely the result of the water/hydroxide equilibrium at the heme iron. It is also possible that changes in the hydrogen bonding of the proximal histidine may also have an influence on the pH dependent redox properties.

Most natural peroxidases have an acidic pH optimum, as having OH^- as the axial ligand makes it more difficult to bind and activate H_2O_2 [152]. The protonation of compound 0, I and II is also aided by a proton rich environment, giving natural peroxidases a preference for operating under acidic conditions. Studies with a thermophilic P450 from *Sulfolobus tokodaii* (P450st) show a pH optimum for peroxygenase activity at mildly alkaline pH's. Experiments with P450st showed that the Lewis acidity of the heme is related to the pK_a^{AA} (H_2O to OH- axial ligand transition) [291]. A mutational study of P450st demonstrated that decreasing the pK_a^{AA} from 8.7 in the WT to 8.1 in the mutant results in increasing the Lewis acidity of the heme and increasing the peroxygenase activity [292]. The k_{cat} of H_2O_2 turnover in the P450st mutant increased sigmoidally with increasing pH, fitting well to a Henderson–Hasselbalch equation. The apparent acid dissociation constant (pK_a^{app}) from this fit was pH 7. This reveals that the rate-limiting step is a deprotonation event with a pK_a of 7, before the formation of the active oxidant [292]. Within the substrate regime being used, this observation could only be attributed to the deprotonation of Compound 0 (Fe^{3+}-H_2O_2). Deprotonation of H_2O_2 before heme binding was not considered as H_2O_2 in solution has a very high pK_a (near pH 12).

With this information in mind we can hypothesise that the observed increase of k_{cat} with increasing pH up to 8.6 is a result of increasing the rate of deprotonation of compound 0. This still leaves the question of why k_{cat} then decreases after pH 8.6? The MD model of C45 shows that the amine groups of 4 lysine residues (29, 44, 52 and 101) are within 15 Å of the heme. In particular, K52 is seen to form a hydrogen bond with both heme propionates and is only around 6.5 Å from the heme edge. The pK_a of these residues is predicted to be between 9 and 11 and changes in their protonation states will likely influence the redox

properties of the heme. The deprotonation of this lysine at alkaline pH could explain the decrease in k_{cat} after pH 8.6 if it is involved in the modulation of the heme redox potential or if it plays a role in protonation the compound 0. At high pH, K52 will be deprotonated and would no longer be able to stabilise the development of negative charge on the bound peroxide during O-O bond cleavage.

2.4.3 Proposed mechanism

Incorporating and activating heme has allowed C45 to access the same redox active species as natural heme oxygenases and peroxidases, namely compound I and II (seen in Figure 2.4.1). These natural peroxidases tend have a conserved structure around the distal site of the heme. HRP, CcP and AXP all have conserved distal histidine and arginine residues. These have been proposed to facilitate the binding of H_2O_2 and then to stabilise the negative charge produced during H_2O_2 cleavage. In the latter stages of the mechanism these residues help facilitate the protonation of compound I and then compound II [119]. A description of natural peroxidase mechanisms can be seen in Section 1.8. C45 does not have this conserved structure of the distal site. We can then conclude that either these structures are not essential or that in C45 other residues are performing and analogous function. The candidate residues that could assist in proton delivery and charge separation are glutamine 45 and lysine 52. The nitrogen of the amide group on Gln45 can be seen as close as 5.8 Å from the heme iron and the amine group of K 52 can be seen 8.3 Å from the heme iron. The second pK_a of the pH dependence of k_{cat}* for C45 is pH 9, making K52 the most likely candidate. The backbone amino groups of leucine 50 and alanine 49 are directly over the distal site of the heme and might also be able to assist in accommodating charged species.

As discussed in Section 1.8.2, some peroxidases, such as CcP, form a compound I where the radical cation is on a nearby tryptophan residue rather than the porphyrin ring itself. C45 has three tryptophan residues, one predicted to be around 7 Å from the heme edge and two on the far end the helix. EPR spectral analysis of C45 indicated that it can form a tryptophan compound I radical. The next chapter will detail the work performed to assign the radical intermediate of C45.

Figure 2.4.1 shows a summary of the C45's peroxide cycle. In the resting state, it appears that C45 has water bound in the distal axial position with a pK_a of 8.1. Without the distal residues found in most natural peroxidases, C45 must wait for spontaneous

deprotonation of the peroxyferric complex. This is inferred from C45's weakly alkali pH optimum, which is similar to P450 peroxygenase activity [292]. The rate of k_{cat} increases as pH increases up to the pK_a of deprotonation of the peroxyferric complex, around pH 8. The rate is then limited at higher pH by the deprotonation of a nearby lysine residue (K52) with a proposed pK_a of 9. This lysine must be involved in the protonation of the heme reactive species, stabilisation of peroxide cleavage, modulation of the redox potential of the heme or some combination of all three.

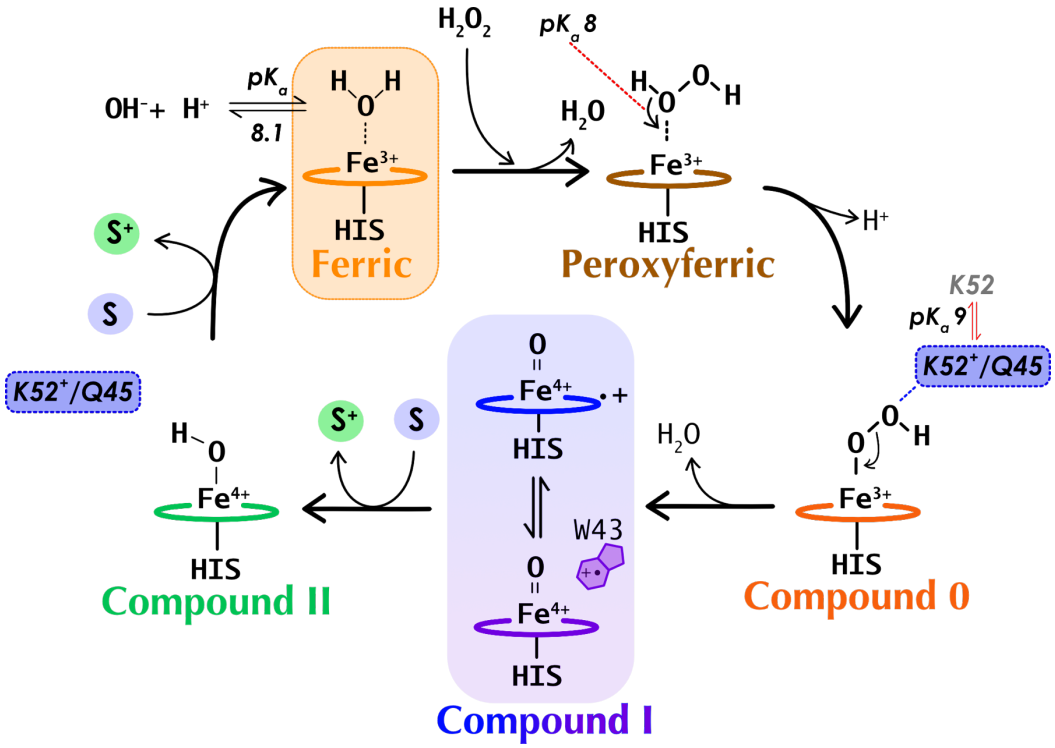


Figure 2.4.1 C45 Mechanism. *A suggested mechanism for the peroxidase cycle of C45. The pK_a values shown in the ferric, peroxyferric and compound 0 are suggested from the interpretation of spectroscopic and kinetic experiments at various pH's*

2.4.4 The sticky question of binding

C45 like many natural peroxidases exhibits activity with a range of electron donating substrates. The steady-state kinetic results for quite different substrates are strikingly similar in terms of the K_M (Figure 2.3.30). This could be because no attempt has been made to design in a substrate binding site. Additionally C45 is small and the heme is relatively exposed. A full ping-pong kinetic analysis was only performed for ABTS with limiting peroxide concentrations were used for all other substrates. If a full ping-pong kinetic analysis could be

calculated for more substrates it could possibly reveal that similarity in observed K_M values could be result of the rate-determining step being linked to another part of the mechanism.

The results of the binding study, that used a computational model of C45 to search for binding sites, were broadly in line with the experimental results. All the substrates that have been kinetically characterised display one or more possible binding loci that persist through 5 ns of molecular dynamics simulation. Homovanillic acid is a possible outlier in that the computational study showed very poor association with the protein, however, preliminary experimental data (Figure 2.3.28) suggests it can be oxidised. These data come with the caveat that Homovanillic acid was not particularly stable in the reaction buffer.

The results of the BUDE docking calculation show several small molecule binding sites on C45. The stability of substrates bound in the small phenolic (SP) pocket within cluster 3, as well as its proximity to the heme, make this a convincing candidate binding site. The fact that this pocket does not appear to be highly selective is in line with very similar experimentally derived K_M values for different substrates. To validate these computational results this pocket should be the target for further computational and experimental studies, producing mutants that perturb the binding in this site. From the frequency of close contact analysis residues L22, E19 and 47 are the 3 top candidates for mutation.

Clusters 1 and 2, from the computational screen, make direct contact with the heme edge. They also appear to be relatively unselective with 10 different substrates binding in these clusters. The proximity to the heme edge means that electron transfer would be rapid and even highly transient interactions with electron donating substrates could result in successful electron transfer. This close but transient interaction could also explain the lack of variation in experimentally derived K_M values. As discussed in Section 1.7.1 most natural peroxidases are generally thought to bind small molecule substrates near the heme edge.

Isoniazid binds to the heme edge in CcP and AXP and has been used to understand the location of binding with other substrates by examining the inhibition it causes [163]. Looking at the inhibitory effects of C45 substrates on the oxidation of a different substrate could be used to study whether they share binding sites. Veratrole alcohol and 1,4-dimethoxybenzene are both seen to bind in the small phenolic pocket of cluster 3 but are not oxidised because of their high redox potentials. These could be the first candidates for use in inhibitory kinetic studies of other substrates that bind in the same pocket, such as TCP.

If one or more of the tryptophan residues form protein radicals, electron transfer could proceed through an alternative pathway, analogous to LiP and CcP. LiP is known to form a tryptophan radical to react with veratrole alcohol. Removing this residue eliminates activity towards veratrole alcohol but electron transfer *via* the heme edge to low potential peroxidase substrates can still proceed. This is especially true in mutants where the normally narrow heme cavity of LiP has been widened [293]. Many natural peroxidases do not have a single well-defined binding site for small molecule peroxidase substrates. CcP, LiP, VP, DypB and DHP have all been seen to have at least two small molecule binding sites with one normally situated near the heme edge [163, 165, 166, 172, 184]. This is also true for C45, the computational study shows 4 well sampled binding sites one of which is at the heme edge. Having several binding sites and the potentially transient nature of interaction at the solvent exposed heme edge could explain the lack of variation in K_M vales from different substrates.

The unifying characteristic that appears to predict substrate reactivity with C45 is the substrate's redox potential. By testing a range of substrates, it is possible to infer that the redox potential of C45 compound I is close to that of 4-aminobenzoic acid (~1000 mV vs SHE). It is important to note that some of the literature midpoint potentials quoted in Table 2.3.5 may have pH dependent redox properties and many not necessarily be the same under the reaction conditions in this project.

The work in this chapter to biophysically, computationally and kinetically characterise C45 has significantly improved our understanding of how this catalytically efficient *de novo* peroxidase functions. Structural characterisation has highlighted that C45 is an ordered but highly flexible protein, which may be a blessing for catalysis but a curse for obtaining atomistic structural information. C45 is structurally and catalytically resistant to high temperatures. The pH dependent k_{cat}* change observed, along with MD simulations, have allowed a detailed catalytic mechanism to be proposed and investigations with many different peroxidase substrates has allowed the redox potential of C45 compound I to be assigned to ~1000 mV vs the NHE. The knowledge developed though these experiments can now go on to inform further design change to future *de novo* enzymes.

Chapter 3

Studying the *de novo* peroxidase radical

3.1 Introduction

As discussed in Section 1.8, some natural peroxidases can form a protein compound I radical. Cytochrome *c* peroxidase (CcP), lignin peroxidase (LiP) and dehaloperoxidase can form tryptophan or tyrosine protein radicals during their reactive cycles [156, 172, 183]. The radical produced by a peroxidase can have a dramatic effect on its activity. In CcP, W191 radical formation is essential for efficient inter-protein electron transfer [207]. In LiP, W171 is required for turnover of certain substrates, such as veratrole alcohol, which cannot be oxidised by many other peroxidases (i.e. HRP, CcP and AXP) [172]. In the *de novo* peroxidase maquette, C45, it is useful to understand what radical intermediate is being produced to test how it affects its activity. The maquettes also offer a unique platform for studying radical formation. Specifically designed radical species could be an important tool for achieving the desired chemistry in future *de novo* enzymes.

For a protein of only 294 residues, CcP has a large number of redox-active amino acids: seven tryptophan and fourteen tyrosine residues [145]. Other peroxidases of a similar size have far fewer redox-active amino acids. Ascorbate peroxidase has two tryptophan and seven tyrosine residues [147]. Lignin peroxidase has only three tryptophan residues [148]. W191 in CcP is required for it to efficiently oxidise its native electron donor ferrocytochrome *c* [207]. Many of the remaining redox-active residues in CcP are capable of radical formation in the absence of an electron donating substrate [216]. These are thought to be involved in the role of CcP in conveying oxidative stress signals, which has been studied in *Saccharomyces cerevisiae* [294].

3.1. Introduction

The role of tryptophan 51 and 191 as well as tyrosine 36, 39, 42, 187, 229, and 236 in CcP have been investigated by mutagenesis [295]. Sequential mutation of these residues to phenylalanine caused an increase in the stability of compound I. The rate of decay of the compound I reduced from 57 s^{-1}, in single W191F mutant, to 7 s^{-1} in the mutant with all seven residues mutated. The characteristic spectroscopic peaks of the compound I were also better defined in the W51,191F mutant [295]. All the mutants displayed very little activity towards ferrocyochrome *c* after the mutation of W191. However, an increase in catalytic efficiency (k_{cat}/K_M) was observed with a small molecule electron donor (2,6-Dimethoxyphenol) as more redox–active residues were removed [295].

The *de novo* peroxidase maquette C45 discussed in the previous chapter contains three tryptophan residues, seen in Figure 3.1.1. The focus of this chapter is to investigate the nature of the radical produced during its reactive cycle [70]. To do this, a series of tryptophan knockout mutants (dW) were designed and characterised (shown in bold in Table 3.1.1).

Protein	Amino acid at position			
	8	43	46	80
C46	**W**	**W**	H	**W**
C45 WT	**W**	**W**	F	**W**
dW3 FFF, W8,43,80F	F	F	F	F
dW3Y FFYF, W8,43,80F F46Y	F	F	Y	F
dW2.1 WFF, W43,80F	**W**	F	F	F
dW2.2 FWF, W8,80F	F	**W**	F	F
dW2.3 FFW, W8,43F	F	F	F	**W**

Table 3.1.1 Tryptophan knock-out mutants.
The protein names in bold are investigated in this chapter.

Two of the tryptophan residues (W8 and W80) are on the on the far end of the four-helix bundle, around 26 Å from the heme group. W43 is on the second helix situated close (around 6 Å) to the heme. One of the proteins investigated (dW3Y) also has a F46Y mutation. F46 is positioned directly above the heme and was the site of the histidine residue in the bis-histidine ligated parent maquette C46 (see Table 3.1.1 and Figure 3.1.1)

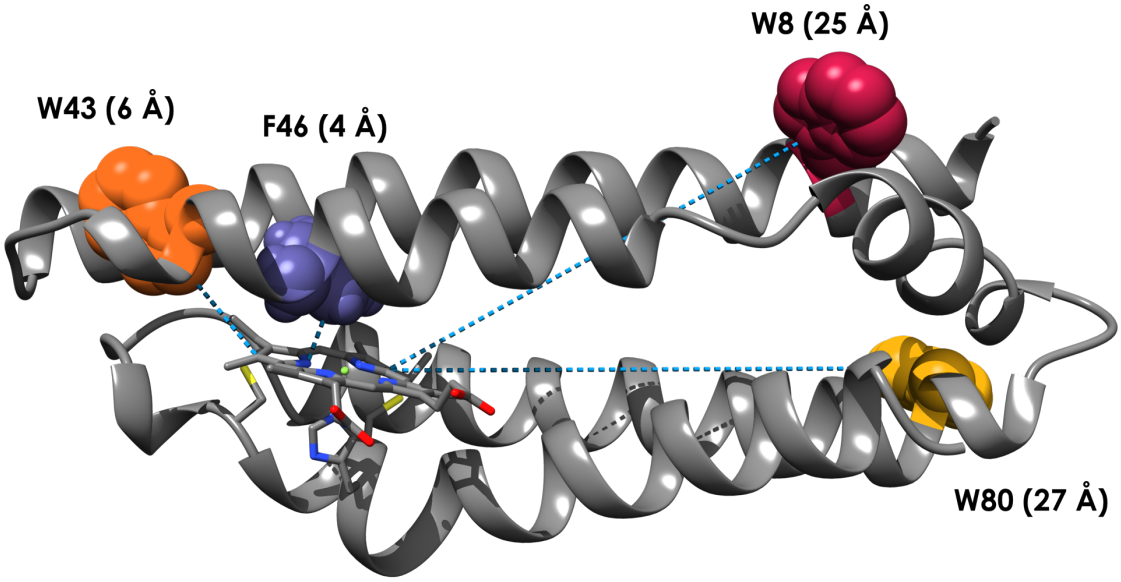


Figure 3.1.1 Tryptophan positions in C45 model.
The residues that have been mutated in this chapter are shown by the Van der Waals radius of their heavy atoms. The closest distance between heavy atoms of the conjugated system of the heme and amino acid side chain are shown by the blue dashed line. The distance to the heme is shown next to the residue label in each case.

3.1.1 Radical scavenging

Radical scavenging is a method of identifying radical species by forming a stable adduct with another molecule [296]. Unlike spin-trapping, where a spin-adduct is formed that retains the radical spin, the adducts formed in radical scavenging are covalent and do not have unpaired spin. TEMPO (2,2,6,6-tetramethylpiperidinyl-1-oxy) is a stable nitroxide radical that is known to scavenge carbon-centred radicals [297]. It has been used identify protein based radicals in CcP and myoglobin by mass spectrometry [296]. The proteins were mixed with H_2O_2 and TEMPO to generate the protein radicals and TEMPO-radical adducts. The samples were digested into peptide fragments and the products were analysed by electrospray ionization (ESI) mass spectrometry (MS). One TEMPO adduct was identified in myoglobin and nine in CcP [296].

3.2 Materials and Methods

3.2.1 Mutagenesis and Cloning

The same cloning and mutagenesis protocols were used as in Section 2.2.1. Mutagenesis was performed using the C45 plasmid, produced in Section 2.3.1, as a starting template for the first round of site-directed ligase-independent (SLIM) mutagenesis, producing the single tryptophan knockout mutants (dW1.1-3 in Table 3.1.1). The successful single tryptophan mutant plasmids were used as templates for a second round of tryptophan knockouts, using the same primers, to produce the double tryptophan mutants (dW2.1-3). One of the successful tryptophan double mutants (dW2.1) was then used in a third round of mutagenesis to produce the triple tryptophan knockout (dW3). To produce dW3Y an additional starting round of mutagenesis was performed to produce the phenylalanine 46 to tyrosine mutant. This was then used as the starting template for three rounds of mutagenesis.

3.2.2 Protein expression and Purification

The same expression and purification protocol was used as described in Section 2.2.2. 4 L cultures of *E. coli* expressing each mutant were grown and purified.

3.2.3 UV-visible spectroscopy

A Cary60 spectrophotometer (Agilent Technologies) was used to record all spectra unless otherwise stated.

3.2.4 CD

CD melts were performed on all dW mutants with the same protocol as detailed in Section 2.2.4. The refolding "cool" experiments were only conducted on dW3 and dWY.

3.2.5 Molecular dynamics

The MD AMBER protocol is found in Section 2.2.6.2. The starting structures were generated using UCSF Chimera to mutate a model of C45 into single, double and triple tryptophan mutants. The dW3Y mutant was then produced from the dW3 model. For each protein, four 100 ns MD simulations were conducted. Although the protein starting

coordinates for each replicate were identical, the solvation and starting velocities were randomly assigned.

3.2.6 Mass Spectrometry

Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry (MALDI-TOF MS) was used to find the mass of the maquettes. 3 μl of 50 μM protein sample was mixed with 3 μl of matrix solution. A small volume of this mixture (<1 μl) was spotted on to steel target plates and left to crystallise and evaporate. The linear positive ion mode on an UltrafleXtreme mass spectrometer (Bruker Daltonics, UK) was used in the University of Bristol's mass spectrometry facility. Matrix solution was equal parts water and acetonitrile, 0.1 % 2,4,6-trifluoroacetic acid (TFA) and a saturating mass of sinapinic acid.

3.2.7 Electrochemistry

Optically transparent thin-layer electrochemistry (OTTLE) allows the UV-visible properties of different redox species of a sample to be monitored *in situ*. It was used to measure the reversible conversion of ferric to ferrous heme iron species of the maquette proteins. A custom built OTTLE cell was constructed from a flat electron paramagnetic resonance (EPR) quartz cuvette (Wilmad, USA). Potentials were applied to samples with a platinum working and counter electrodes from a Biologic SP-150 potentiostat. An $AgCl_2$ reference electrode (BASi, USA) was used. All recorded potentials have been converted to mV vs the Nernst hydrogen electrode (NHE) by the addition of 270 mV, determined through calibration with a known protein sample. The following mediators were added (potentials are given in mV vs NHE): 20 µM methyl viologen (-430 mV), 20 µM benzyl viologen (-311 mV), 20 µM anthroquinone-2-sulfonate (-225 mV), 20 µM phenazine (-180 mV), 25 µM 2-hydroxy-1,4-napthoquinone (-152 mV), 6 µM indigotrisulfonate (-90 mV), 10 µM pyocyanine (-34 mV), 50 µM duroquinone (5 mV), 20 µM phenazine ethosulfate (50 mV), 20 µM phenazine methosulfate (80 mV) and 25 µM 1,2-napthoquinone (130 mV) .

120 μl of 50 μM protein samples were prepared with 10% glycerol and mediators in redox buffer (100 mM KCl, 20 mM CHES at pH 8.6). This was then pipetted into the cell. The working electrode was then submerged into the sample and any bubbles gently removed. Next, redox buffer without glycerol was pipetted on top of the sample taking great care not to disrupt the interface between sample and buffer. The counter and reference electrodes were added to the buffer only layer of the cell. Increasingly reducing potentials were applied to the sample followed by increasingly oxidising potentials. Each potential was held for 30 minutes to ensure that the cell was well equilibrated. A baseline run was taken in the same conditions but without the addition of protein to the sample solution. The baseline readings were subtracted from the protein measurements at each potential.

The ferrous Soret λ_{max} showed the highest dynamic range in absorbance during the experiments. It was used to normalise the data to the fraction of reduced protein throughout to the experiment. This was plotted against the potential and fitted to either a single (3.1) or 2 x 1 electron Nernst equation (3.2):

$$R = \frac{A+B.10^{\left(\frac{(E_m-mV)}{59}\right)}}{1+10^{\left(\frac{(E_m-mV)}{59}\right)}} \tag{3.1}$$

$$R = \frac{\left(A.10^{\left(\frac{(E_m1-mV)}{59}\right)}+B+C.10^{\left(\frac{(E_m2-mV)}{59}\right)}\right)}{\left(1+10^{\left(\frac{(E_m1-mV)}{59}\right)}+10^{\left(\frac{(E_m2-mV)}{59}\right)}\right)} \tag{3.2}$$

"mV" is an independent variable referring to the applied potential vs NHE. "R" is the fraction of ferrous heme iron. "A" and "B" are coefficients corresponding to the maximum and minimum values of R, respectively. "Em" refers to the midpoint potential of heme iron.

3.2.8 Steady-state kinetics assays

Michaelis–Menten kinetic data were obtained using a KinetAsyst SF-61DX2 Double-mixing Stopped-Flow Spectrophotometer (Hi-Tech Limited, UK) with a 1 cm path length. Reactions were initiated by mixing equal volumes of one solution containing 0.1 μM protein and various concentrations of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and another solution of 100 μM H_2O_2. Both solutions were buffered with redox buffer

at pH 8.6. The initial linear portion of each trace was fit to a straight line and the gradient (ΔAbs.Sec^{-1}) determined using an interactive python script. The initial ranges for each substrate concentration were then fit to a Michaelis–Menten curve using GraphPad Prism. Turnover of ABTS was monitored at 405 nm (ε405 nm = 36,800 M^{-1}.cm^{-1}) [250]. The ABTS concentrations used were 1, 5, 10, 20, 40, 80, 100, and 150 μM.

3.2.9 Total Turnover (TON)

Turnover number (TON) is the number of moles of substrate that can be turned over per mole of enzyme. In an enzyme that is not inhibited by its substrate or product, this can be calculated by reacting a small enzyme concentration in an excess of substrate. High peroxide concentrations cause destruction of the heme in C45. The high K_M for peroxide in C45 means that at lower peroxide concentrations reactions proceed very slowly. This combination of a high K_M and heme degradation made choosing reaction conditions difficult. Brief tests were conducted with guaiacol, 2,4,6-Trichlorophenol (TCP) and ABTS as electron donating substrates, and glucose oxidase (GOX) and glucose to produce a steady supply of peroxide. However, this setup was not taken further as it requires extensive optimisation to produce a reliable assay.

TON assays for all the dW mutants and the WT were conducted in a 96-well format on a Synergy Neo2 Plate Reader (BioTek, UK). Six replicate wells were set up for each protein with 100 μl of ABTS and protein to be at a working concentrations of 2.4 mM and 10 nM respectively. Spectral scans were taken before the injection of 100 μl H_2O_2 to be at a working concentration of 500 μM. Redox buffer at pH 8.6 was used for all samples. ABTS radical cation production was monitored and a final absorbance value taken at 120 min. Two controls were also set up: one an injection of buffer into a sample of C45 and ABTS; the second was injecting H_2O_2 in to ABTS without any enzyme present. This was used to baseline the results from the protein samples. The concentration of ABTS turned over was calculated using the 405 nm extinction coefficient (ε405 nm= 36,800 M^{-1}.cm^{-1}) [250].

3.2.10 Heme degradation assay

Heme degradation was monitored using SX20 Stopped-Flow (Applied photophysics) fitted with a photodiode array detector (PDA). 8 μM of protein was mixed with 800 μM H_2O_2,

both in redox buffer at pH 8.6 at 25 ˚C. The decrease in Soret λ_{max} was monitored at 405 nm; repeated measurements were fit to a two-phase exponential decay using GraphPad Prism.

3.2.11 Intermediate formation

Intermediate formation was monitored using a SX20 Stopped-Flow (Applied photophysics) fitted with a photodiode array detector (PDA). 8 μM of protein was mixed with 800 μM H_2O_2, both in redox buffer at pH 8.6 at 25 ˚C. The increase in 520 nm signal was monitored until it started to decay (~5 seconds). Repeated measurements were fit to a one-phase exponential association using GraphPad Prism.

3.2.12 Radical scavenging

Preliminary radical scavenging experiments were conducted with TEMPO (2,2,6,6-tetramethylpiperidinyl-1-oxy). TEMPO is a stable nitroxide radical that is known to scavenge carbon-centred radicals [297]. Two different MS methods were tested.

The first was using the Proteomics Facility at the University of Bristol. 20 μl reaction samples contained: 3 μM C45 or myoglobin, 200 μM H_2O_2 and 200 μM TEMPO. Controls without H_2O_2 and TEMPO were also prepared. Samples were reacted for 10 minutes before tryptic digest overnight by the Proteomics Facility. Fragments were then analysed by nano-LC MSMS using an Orbitrap Fusion Tribrid mass spectrometer (Thermofisher, UK).

The second method was without tryptic digest and using MALDI-TOF MS. 50 μM protein, 500 μM H_2O_2 and 2000 μM TEMPO were made up in 5 μl reaction volumes. After 10 minutes incubation at room temperature, samples were prepared by the same methods as described in Section 3.2.6.

3.2.13 EPR

Electron Paramagnetic Resonance (EPR) was performed at the Manchester Institute of Biotechnology (The University of Manchester, UK). An ELEXSYS E580 spectrometer (Bruker, UK) was operated by Dr Muralidharan Shanmugam (The University of Manchester) who also gave advice on sample preparation. X-band cw-EPR spectra were recorded of the resting state and radical protein samples at 12 K. A narrow magnetic field sweep of 100 Gauss was used to sample the radical signal. Samples of the resting state were 220 μl of 300 μM protein in redox buffer pH 8.6. These were loaded into quartz EPR sample

tubes and carefully frozen in liquid nitrogen. Radical samples were produced by mixing 220 μl of 300 μM protein with 24 μl of 10 mM peracetic acid (PAA) in redox buffer pH 8.6. The working concentration of PAA was then close to 1 mM. Samples were then rapidly loaded in to sample tubes and frozen in liquid nitrogen. Care was takes to avoid sample tube fracture during freezing. The time taken for mixing, loading and freezing was around 3 seconds.

3.3 Results

3.3.1 Design and mutagenesis

The template used for producing the tryptophan knockout (dW) mutants was the vector containing C45 (the mono-his ligated *c*-type maquette), produced in Section 2.3.1. Primers were designed to remove W8, 43 and 80 replacing them with phenylalanine. Primers for mutating phenylalanine 56 to tyrosine were also produced. Table 3.3.1 shows the sequential mutations made. The dW3 naming convention corresponds to the full tryptophan knockout. The name dW2.1 corresponds to the mutant with two tryptophan residues knocked out with one remaining in the first position. The mutagenesis was determined to have been successful by sequencing the relevant portion of the plasmid (Eurofins, MWG).

Protein	Amino acid at position			
	8	43	46	80
C46	**W**	**W**	H	**W**
C45 WT	**W**	**W**	F	**W**
dW3 W8,43,80F	F	F	F	F
dW3Y W8,43,80F F46Y	F	F	Y	F
dW2.1 W43,80F	**W**	F	F	F
dW2.2 W8,80F	F	**W**	F	F
dW2.3 W8,43F	F	F	F	**W**
dW1.1 W8F	F	**W**	F	**W**
dW1.2 W43F	**W**	F	F	W
dW1.3 W80F	**W**	**W**	F	F
dW1.1Y W8F F46Y	F	**W**	Y	**W**
dW1.2Y W43F F46Y	**W**	F	Y	**W**
dW1.3Y W80F F46Y	**W**	**W**	Y	F
dW2.1Y W43,80F F46Y	**W**	F	Y	F
dW2.2Y W8,80F F46Y	F	**W**	Y	F
dW2.3Y W8,43F F46Y	F	F	Y	**W**

Table 3.3.1 Tryptophan knock-out genes produced

This table details the maquette genes produced by sequential SLIM mutagenesis of C45. The protein names in bold were investigated in this chapter. The genes below the dotted line were not used to express protein but were produced for possible further study.

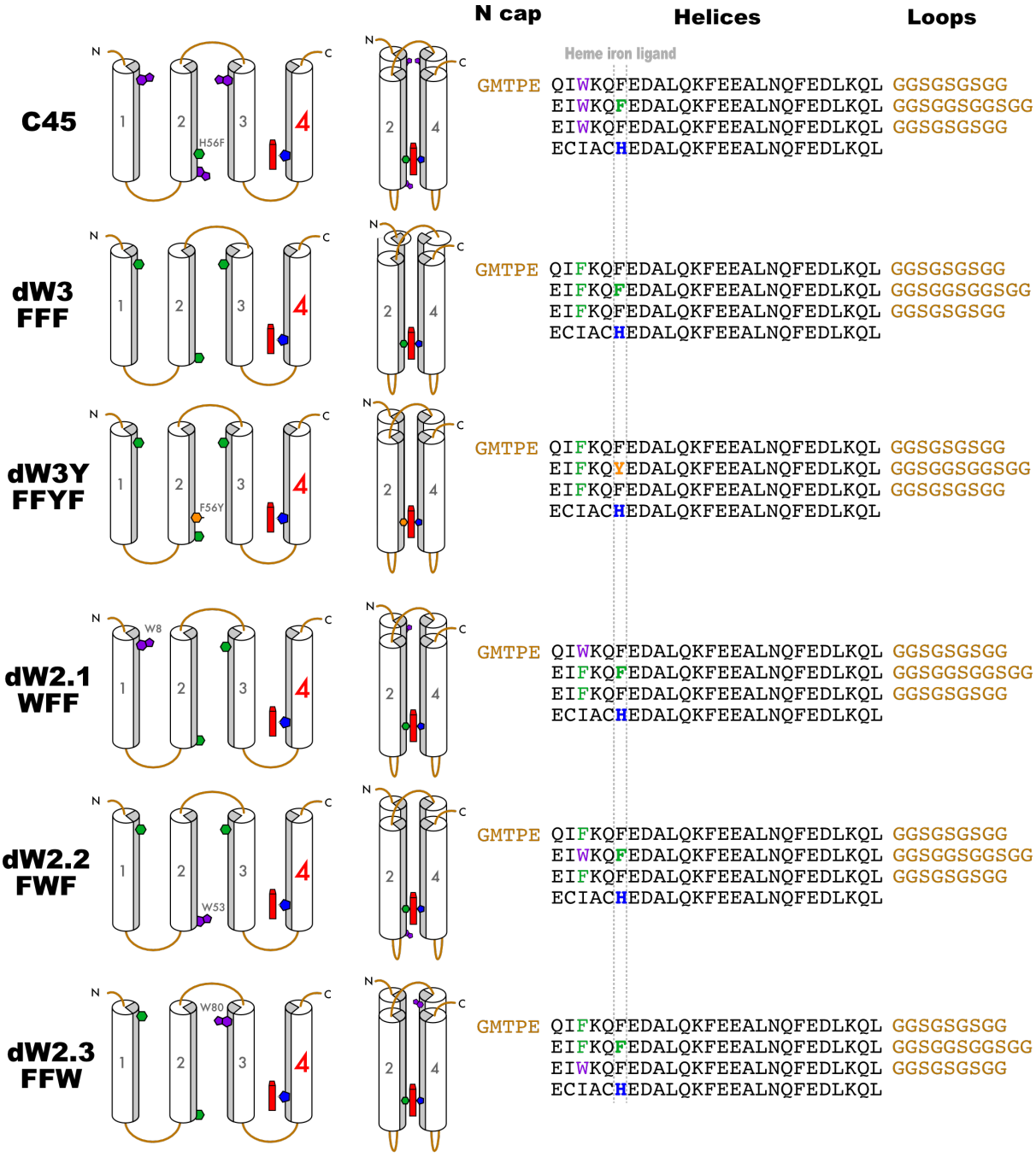


Figure 3.3.1. Design of tryptophan knock-out maquettes

The coloured residues in the protein sequence correspond to the coloured residues displayed both in the loop schematic and the bundle schematic. The cylinders represent the helices of the maquettes with the grey shaded portion depicting hydrophobic residues.

3.3.2 Expression and purification

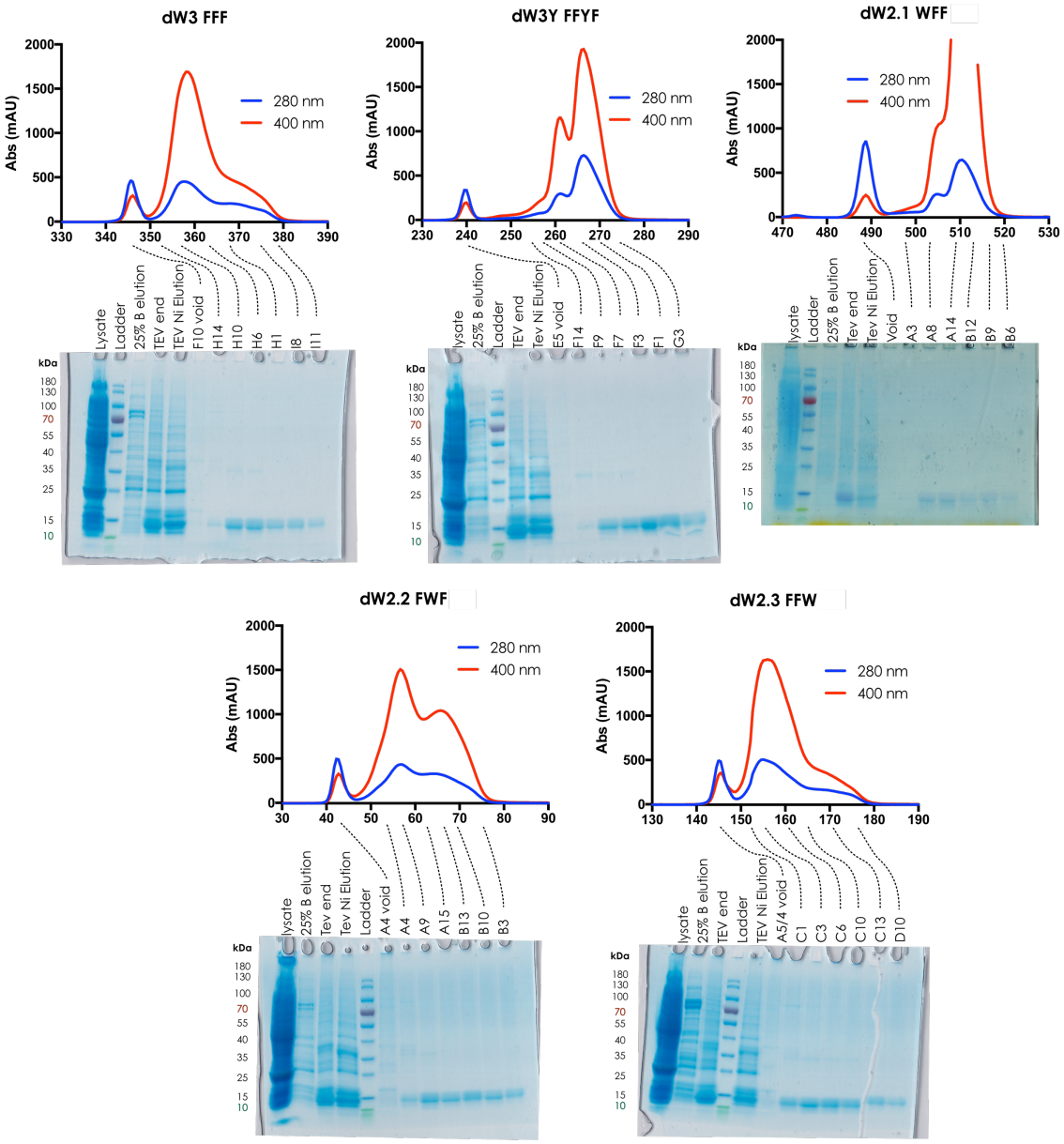


Figure 3.3.2 SEC and SDS-PAGE results for dW mutants.

For each protein the 280 nm and 400 nm SEC trace is shown. The x-axis shows the accumulated volume (in ml). The first peak in each case is the void volume eluting around 43 ml after injection. Below each trace is the SDS-PAGE results of the purification. The bands (not including the ladder) start with the cell lysate then "25% B elution" which is a wash of proteins weakly associated to the first nickel affinity column. The "TEV end" lane is the elution from the first nickel affinity column after the cleavage of the his-tag. The "TEV elution" is the elution from the second nickel column containing contaminants that bind to nickel and uncleaved sample protein. The remaining bands are samples of 3 ml fractions from the SEC purification of the flow-through from the second nickel column.

The expression and purification of the tryptophan knockouts was performed in the same way as C45 in Section 2.2.2. The new maquettes expressed and incorporated heme well. Figure 3.3.2 shows a size exclusion chromatography (SEC) trace and SDS-PAGE results for

each of the mutants during the purification. The SEC traces show that the maquettes are not in a homogeneous conformation and may associate transiently as dimers at pH 8.6.

3.3.3 UV Visible spectroscopy

The UV visible spectrum of the resting state ferric (Fe^{3+}) and oxidised ferrous (Fe^{2+}) heme of all the tryptophan knockout mutants (dW) were recorded. The ferrous spectrum shown in Figure 3.3.3 are taken from OTTLE experiments. Traces were all normalised to the Soret λ_{max} value of the WT protein before conducting the curve fitting. Fityk curve fitting software was used to deconvolute the spectra of the ferrous maquettes [298]. It allowed the intensity of the 430 nm shoulder peak to be quantified (see Figure 3.3.4.B).

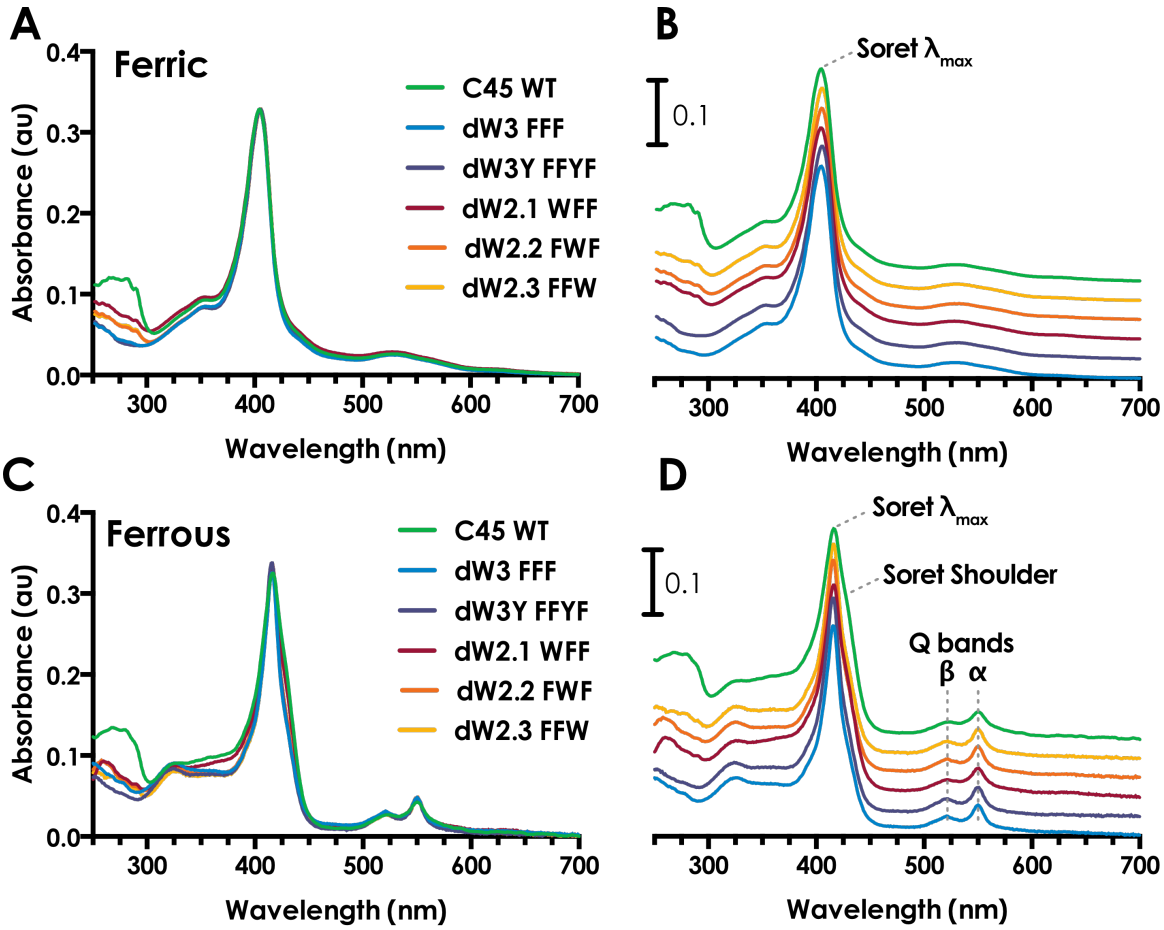


Figure 3.3.3 UV-Visible spectroscopy of dW mutants. *(A) and (B) are the spectra of the maquette in the ferric form. (B) is the same results as (A) in a staggered arrangement. (C) and (D) are the spectra of the reduced maquettes in the ferrous form. (D) is the same results as (C) in a staggered arrangement.*

	Peak	Wavelength (nm)					
		C45	dW3 FFF	dW3Y FFYF	dW2.1 WFF	dW2.2 FWF	dW2.3 FFW
Ferrous	soret 0	412.8	412.6	412.3	412.9	412.3	413.0
	Soret λmax	416.4	415.8	415.5	416.1	415.7	416.0
	Shoulder	430.6	430.4	430.1	430.4	430.2	430.2
	Q-β	521.4	521.8	521.1	521.7	521.8	521.6
	Q-α	550.7	550.4	550.2	550.5	550.4	550.4
Ferric	Soret Max	405.0	405.0	406.0	405.0	405.0	406.0

Table 3.3.2 Deconvoluted UV visible peaks. *Fityk curve fitting software was used to deconvolute the Soret and Q bands. The Soret peak could be fit with just the Soret λ_{max} and 430 nm shoulder peak, but including a third absorbance peak (Soret 0 at 413 nm) resulted in a better fit.*

The positions of the peaks of all the dW mutants are very similar to C45, differing by only 1 nm or less. There is a clear reduction in the 280 nm absorbance in the dW mutants compared to C45 (see Figure 3.3.4.A). This is expected as tryptophan contributes to absorbance at 280 nm There is no observable increase in 275 nm absorbance for the tyrosine mutant (dW3Y).

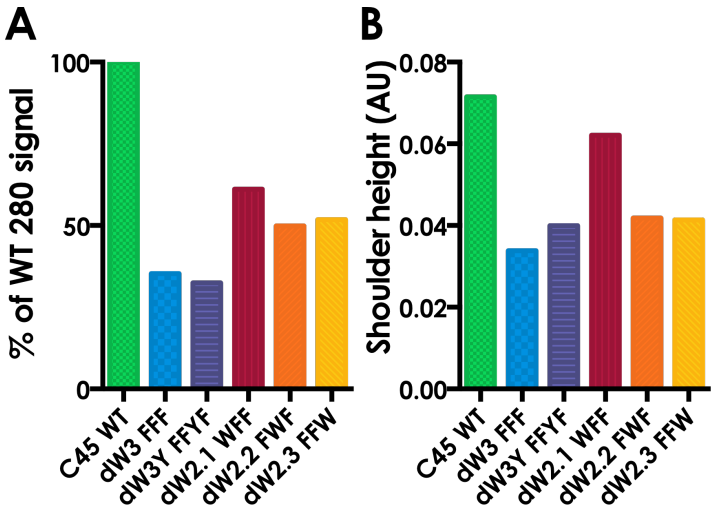
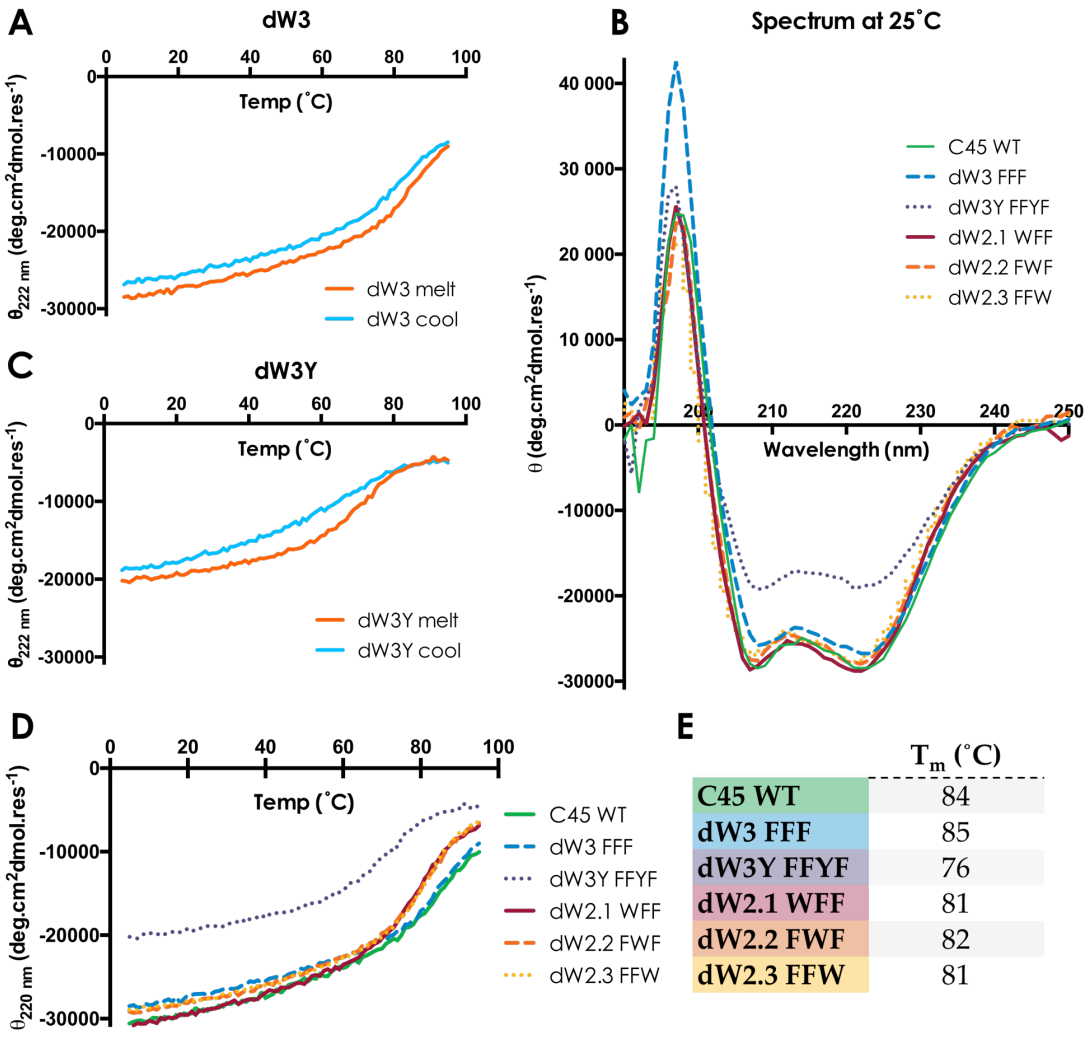


Figure 3.3.4 Ferric 280 nm signal and ferrous 430 nm shoulder heights comparison. *280 nm values for* ***(A)*** *were taken from ferric spectra normalised to the Soret λ_{max} of C45. Values for* ***(B)*** *were derived using Fityk curve fitting software on the normalised ferrous spectra.*

3.3.4 CD

CD spectrum of the WT and dW mutants shows the typical absorbances for α-helical proteins. These are negative bands at 222 nm and 208 nm and a positive band at 193 nm [299]. The positive band at 193 nm for the dW3 mutant is particularly pronounced. dW3Y is the only mutant to show a significant spectral difference in helical signal. dW3Y is less helical and has a lower melting temperature T_M = 76 °C. The melting temperatures for the single tryptophan mutants (dW2.1-3) are all very similar and only 3 or 2° lower than the WT. The refolding experiments conducted with dW3 and dW3Y show refolding but to a lesser extent than observed in C45 WT (Figure 2.3.4).



	T_m (°C)
C45 WT	84
dW3 FFF	85
dW3Y FFYF	76
dW2.1 WFF	81
dW2.2 FWF	82
dW2.3 FFW	81

Figure 3.3.5 CD results comparison for dW mutants.

(A) and (C) show the folding and re-folding of dW3 and dW3Y between 5 and 95 °C, the trace monitors the mean residue ellipticity (θ) at 222 nm. (B) shows a comparison of the CD spectrum of the five dW mutants and the WT at 25 °C. (D) is a comparison of the same 6 proteins unfolding from 5 to 95 °C. Table (E) displays the T_m of unfolding calculated as the temperature with the maximal gradient. All experiments were conducted with 10 μM protein in Redox buffer at pH 8.6.

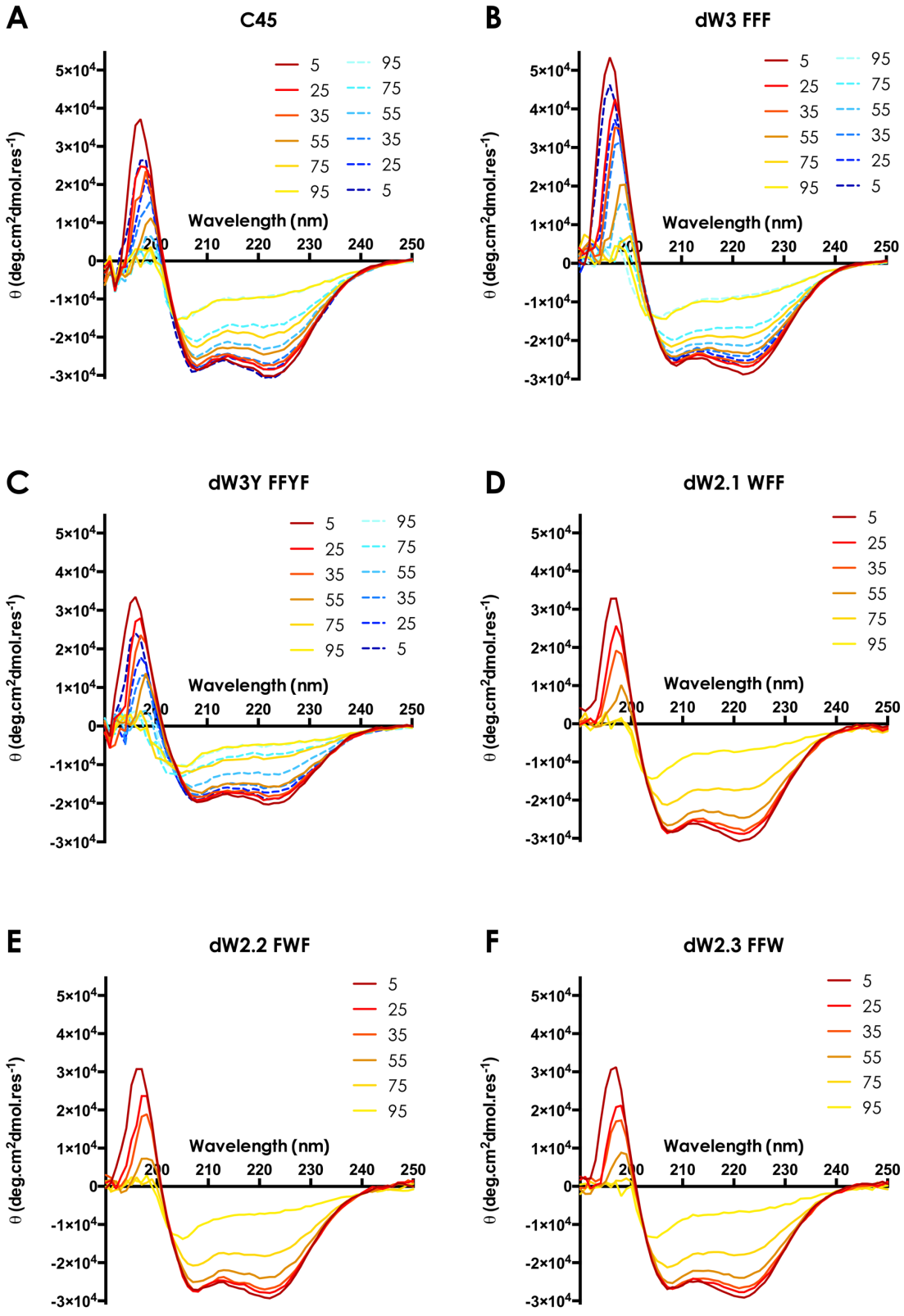
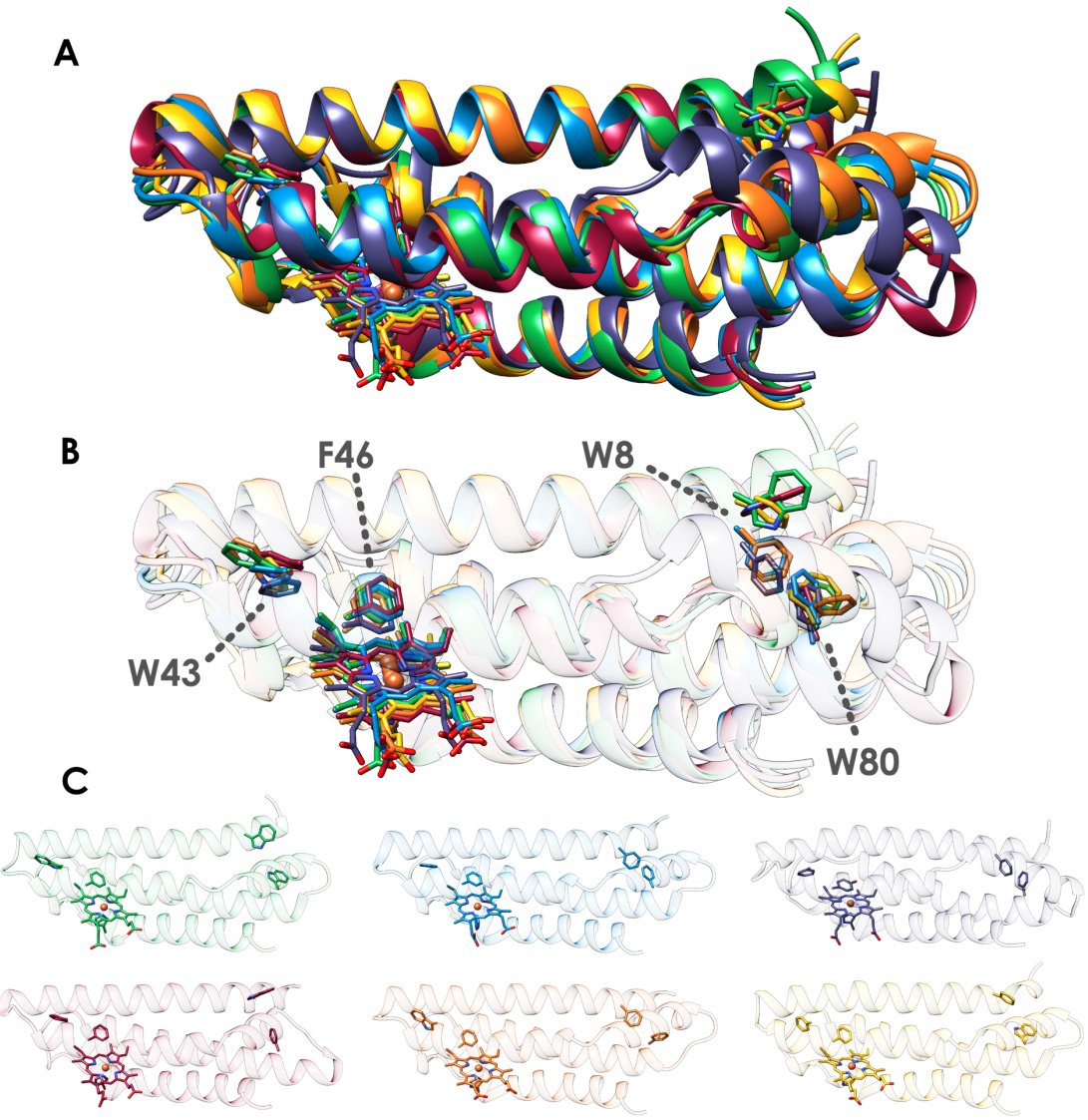


Figure 3.3.6 CD thermal melt spectrum scans for dW mutants and C45. *Scans of melts for 5-95 °C are displayed as solid lines going from red to yellow. Scans of cooling experiments from 95-5 °C are displayed as dashed line from light to dark blue. All experiments were conducted with 10 μM protein in Redox buffer at pH 8.6.*

3.3.5 Molecular dynamics

The MD of each protein was run in four replicates of 100 ns each. Each replicate had unique starting velocities and solvation patterns to increase the level of conformational sampling. The mutations did not cause any large structural changes from the WT C45. Figure 3.3.8.A shows the average and variance of the RMSD in each protein compared to an average backbone structure of all proteins. The RMSD are all near 1.5 Å and stable indicating that they all adopt very similar structures.



Key: C45 WT - dW3 FFF - dW3Y FFYF - dW2.1 WFF - dW2.2 FWF - dW2.3 FFW

Figure 3.3.7 MD structures. *The colour of each structure corresponds to the colour key above.* ***(A)*** *is an alignment of MD structures from each of the 6 maquettes.* ***(B)*** *shows the same viewpoint as (A) with the backbone ribbons made transparent to show the positions of mutated residues more clearly. The labels correspond to the residues in C45 WT.* ***(C)*** *is the structures that make up (B) separated for clarity.*

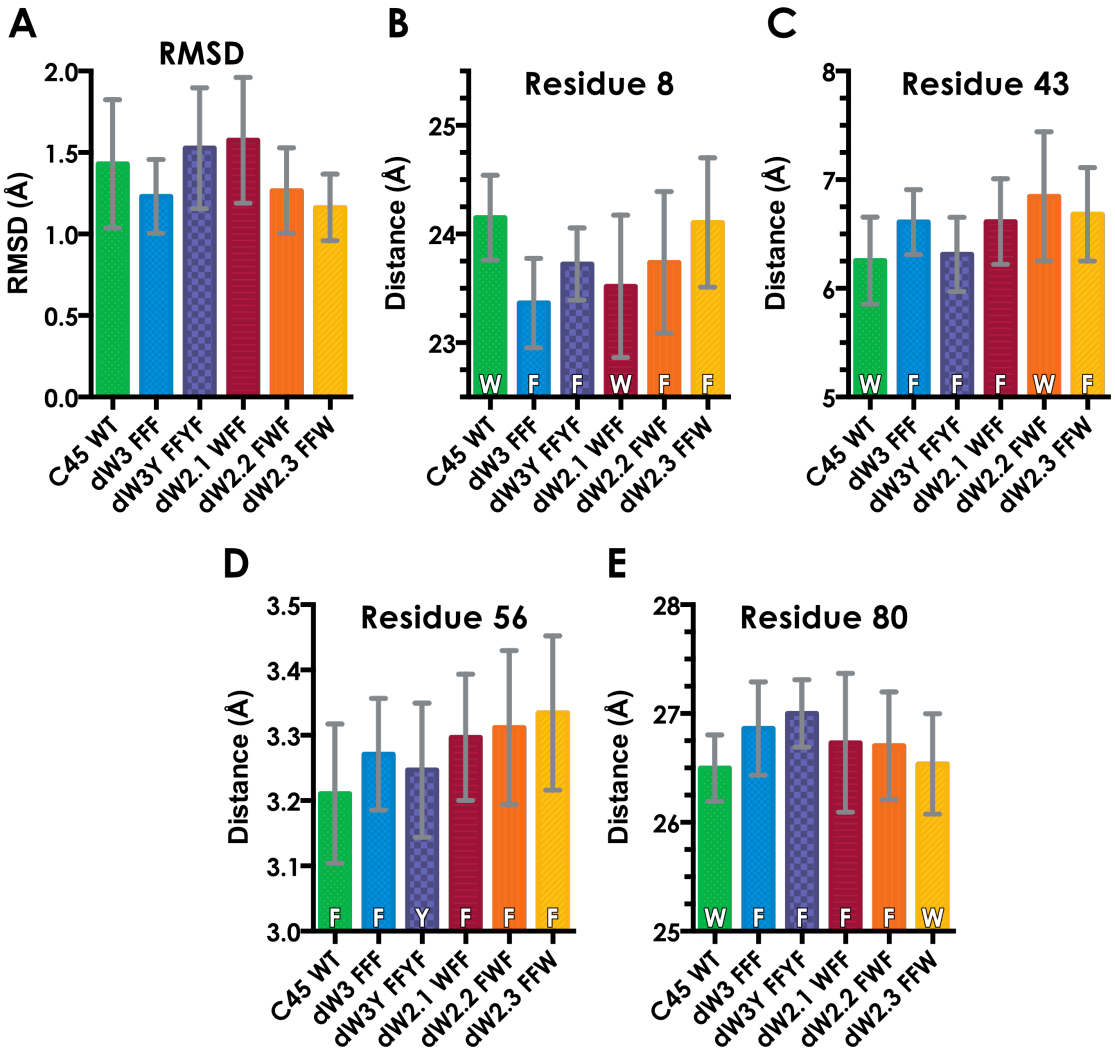


Figure 3.3.8 Averaged MD data. *(A) is the average RMSD from all frames of all four replicates. The RMSD is compared against to an average backbone structure of across all replicates of all mutants.* ***(B-E)*** *shows the shortest distance between the heavy atoms of the heme conjugated system and heavy atoms of the side chain conjugated system for the WT positions; W8, W43, F46 and W80 respectively. The shortest distance was measured at each frame (every 100 ps) and the value graphed is an average across all frames and replicates. The error bars show the standard deviation. The white letters at the bottom of each bar are the residue in that position for that protein.*

The shortest distance between the heme and the side chain was defined as the shortest distance between the heavy atoms of the heme conjugated system and heavy atoms of the side chain conjugated system. These average distances for each residue position deviate by less than 1 Å from each other. The most notable structural change is that the W8F mutation moves to sample a more buried conformation (seen in Figure 3.3.7: dW3, dW3Y and dW2.2 FWF). The mutation W80F also causes the W8 position to sample conformations, on average, slightly closer to the heme as it causes a small movement in the end of the first helix. The

combination of the four mutations in dW3Y has also caused the existing kink in the second helix to grow. The π-stacking interaction seen in C45 WT between F46 and the heme is maintained in all mutants including dW3Y, with the F56Y mutation. The side chains of K52 and Q45, which are likely to play a role in the catalytic cycle of these peroxidase maquettes, are found close to the heme edge in all mutants. K52 samples a conformation that hydrogen bonds to the heme propionates in all mutants with a variance comparable to C45 WT.

3.3.6 Mass Spectrometry

MALDI-TOF MS with a sinapinic acid matrix was performed for the WT and all dW mutants. The laser intensity required was 95 % of its maximum which could be the cause of some fragmentation and mass adducts. All the major mass peaks shown in Table 3.3.3 are heavier than expected. Common MALDI adducts and their mass increases include: hydrogen (1), oxygen oxidation (16.0), water (18.0), sodium (22.8), potassium (39.0) and methyl cyanide (42.0). The protein sample buffer contained potassium ions. The inclusion of a potassium adduct and the high laser strength causing degradation or fragmentation could account for the variability in the observed mass. Figure 3.3.9 shows the major mass peaks in the region 14000 to 16500 m/z. Lower mass peaks, corresponding to the doubly ionised analyte, were seen in all samples. In the samples with higher signal intensities, very small peaks at double the mass of the major peak could be seen, corresponding to protein dimers.

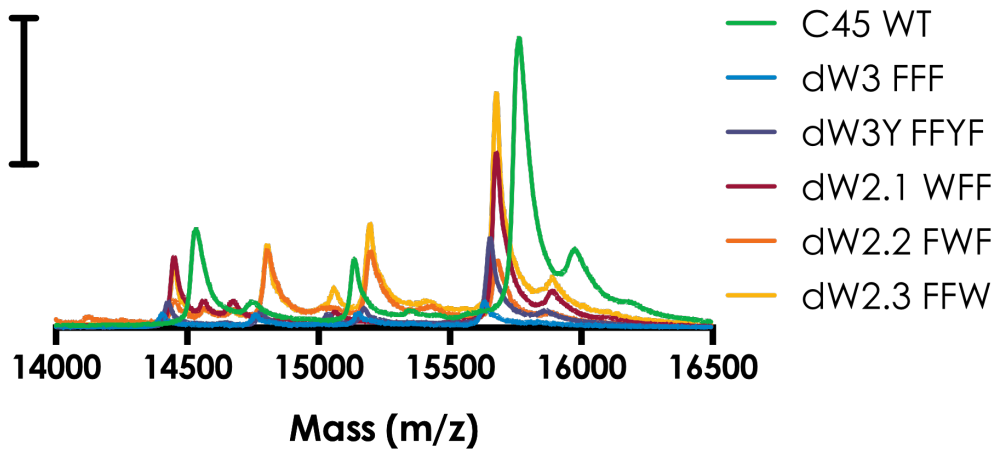


Figure 3.3.9 Mass spectra. *The peaks show an overlay of the mass peaks for all samples. The y axis is the intensity in arbitrary units. The scale bar is 20,000 (AU). The spectra have been clipped to show the major peaks of the samples. Lower mass peaks were present in all samples for the doubly ionised analyte.*

Comparing the difference between the WT and mutant masses shows mass changes consistent with the mutations being successful. The greatest deviation in mass from the WT was for the dW3 mutant which has three tryptophan residues removed. The mutants with two tryptophan residues removed (dW2.1-3) are all a similar mass, closer than the other mutants to the WT mass.

	Mass (m/z)					
	C45 WT	**dW3 FFF**	**dW3Y FFYF**	**dW2.1 WFF**	**dW2.2 FWF**	**dW2.3 FFW**
MALDI Peak	15763.57	15630.65	15651.62	15674.67	15679.69	15676.72
MALDI -WT	0.00	-132.92	-111.95	-88.91	-83.88	-86.85
Predicted	15789.05	15671.94	15687.94	15710.98	15710.98	15710.98
Predicted - WT	0.00	-117.11	-101.11	-78.07	-78.07	-78.07
MADI - Predicted	25.48	41.29	36.31	36.31	31.29	34.26

Table 3.3.3 MADLI mass peak compared to predicted mass. *The first row shows the mass of the major peak from MALDI TOF MS. The second row shows the difference between the WT mass peak and the other samples. The third and fourth rows show the predicted mass of the protein with heme and the difference from the WT protein expected with each mutation. The last row shows the difference between the experimental mass and predicted mass.*

3.3.7 Electrochemistry (OTTLE)

Figure 3.3.10 shows the midpoint potentials fit from spectra taken during the OTTLE experiments. The midpoint potentials of dW3, dW3Y and dW2.3 FFW are all quite similar, between 6 and 7 mV more positive than the WT. The midpoint potential for dW2.2 FWF and dW2.1 WFF mutants are 10 and 15 mV more positive than the wild type. Changes in the heme environment lead to changes in the midpoint potential of the heme. The changes in midpoint potential do not seem additive with removal of two and then three of the tryptophan residues. This would suggest that the changes are down to subtle structural rearrangements associated with adding more hydrophobic residues to the core of the protein.

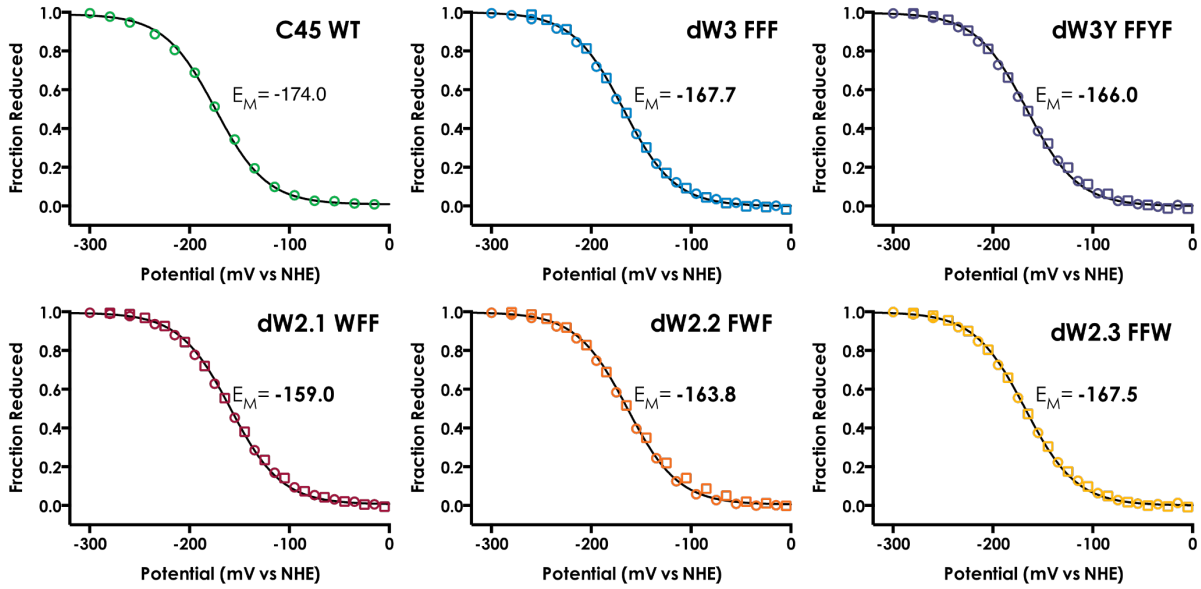


Figure 3.3.10 OTTLE results. *The fraction of reduced protein has been fitted to a single electron Nernst equation. The midpoint potentials (E_M) are shown on each graph. The values taken during the reduction are shown in circles and the subsequent oxidation values are shown as squares. All EM values from the curve fitting have a standard deviation of less than 1.*

3.3.8 Steady-state kinetics with ABTS

The steady-state kinetics for the tryptophan mutants was performed using ABTS as an electron-donating substrate and with 100 μM H_2O_2. The k_{cat} values for the different mutants remained similar to the WT. The exception to this was the dW2.2 FWF mutant which showed a 0.8 s^{-1} reduction from the WT. All the dW mutants show a decrease in K_M which, in classical Michaelis–Menten kinetics, would indicate that the enzyme requires a lower concentration of substrate to become saturated. Including the tyrosine at position 46 in the dW3Y mutant increases K_M. There is also a slight increase in K_M moving from the dW3 mutant to those with a single tryptophan remaining, the greatest increase coming from the dW2.2 FWF mutant where the tryptophan is in the position close to the heme. As a result of the decrease in K_M, the k_{cat}/K_M for dW3, dW2.1 WFF and dw2.3 FFW have more than doubled.

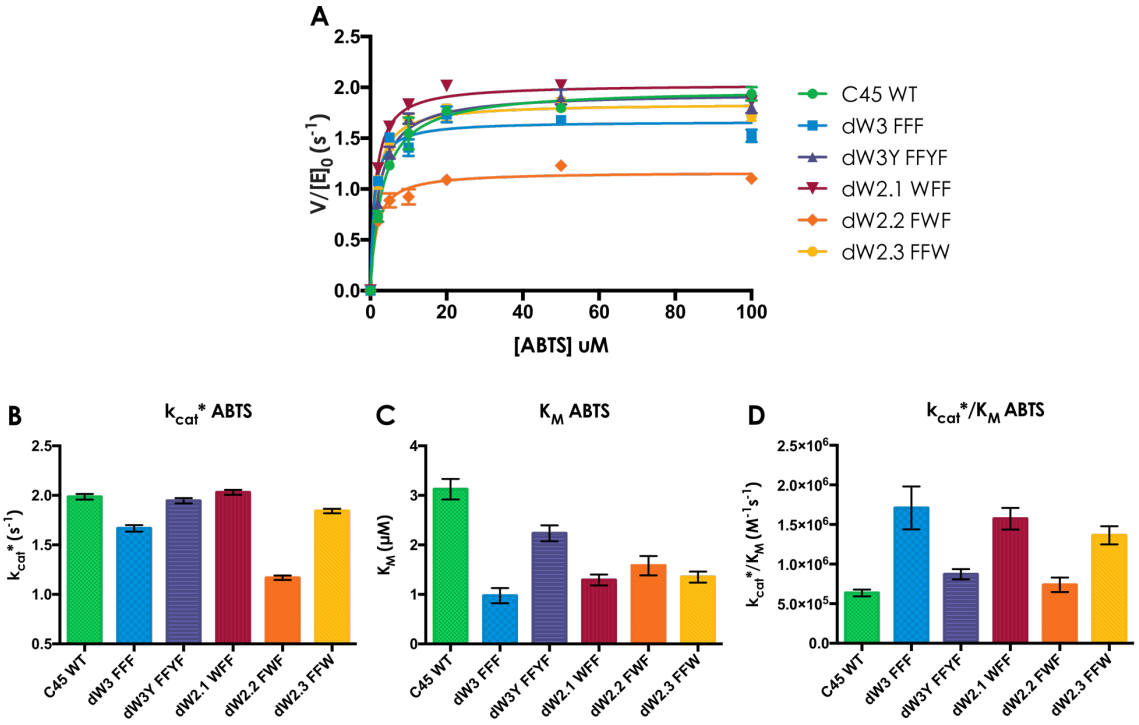


Figure 3.3.11 Steady-state results for dW mutants.

(A) Steady-state kinetics at (rate-limiting) 100 μM H_2O_2, fitted to a Michaelis–Menten curve. (B) shows the k_{cat} values at rate limiting H_2O_2. (C) shows the K_M values. (D) shows the k_{cat}*/K_M values. Error bars represent the standard deviation.*

Enzyme	Substrate	k_{cat}* (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	Condition
C45 WT	ABTS	2.0* ± 0.028	3.10 ± 0.21	6.4 ± 0.43 x10 ^{5}	pH 8.6
dW3 FFF	ABTS	1.7* ± 0.033	0.98 ± 0.15	1.7 ± 0.27 x10 ^{6}	pH 8.6
dW3Y FFYF	ABTS	1.9* ± 0.026	2.20 ± 0.16	8.7 ± 0.64 x10 ^{5}	pH 8.6
dW2.1 WFF	ABTS	2.0* ± 0.025	1.30 ± 0.11	1.6 ± 0.14 x10 ^{6}	pH 8.6
dW2.2 FWF	ABTS	1.2* ± 0.023	1.60 ± 0.19	7.4 ± 0.92 x10 ^{5}	pH 8.6
dW2.3 FFW	ABTS	1.8* ± 0.023	1.40 ± 0.11	1.4 ± 0.11 x10 ^{6}	pH 8.6

Table 3.3.4 Comparison of dW mutant kinetic results.

*The Michaelis–Menten kinetic fit results for ABTS with the dW mutants and WT (column 1) have been tabulated. The * result is that k_{cat}* has been obtained at limiting H_2O_2 with 100 nM C45 and 100 μM H_2O_2. The ± errors quoted are standard deviation.*

3.3.9 Turnover Number (TON)

The turnover number (TON) is a measure of the number of moles of substrate a mole of enzyme can turnover. High experimental variation and relatively few repeats (n=6) meant that only comparisons of the WT to dW3Y and WT to dW2.2 are statistically different (using a Tukey's multiple comparisons test). If the trend of this result was extended over further repeats it would suggest that retaining the tryptophan has a protective effect on the C45 WT and that the effect comes from the tryptophan residues far from heme.

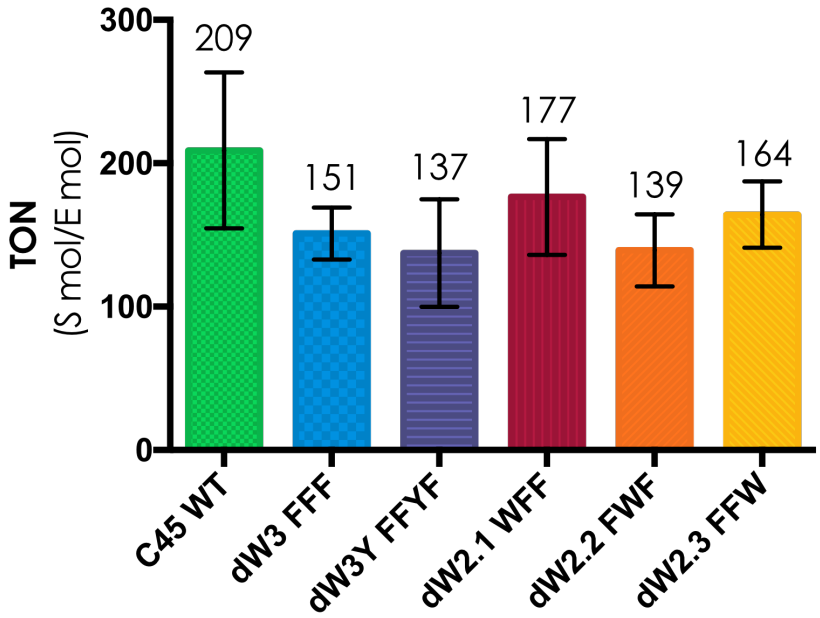


Figure 3.3.12 Turnover number dW mutants with ABTS.
Means are displayed above each bar. Error bars shown are the standard deviation.

3.3.10 Heme degradation assay

The fast rate of the two-phase exponential decay corresponds to the formation of the reactive intermediate. The slow rate is the degradation of the heme by the reactive intermediates in the absence of an electron donating substrate. dW3 and dW2.1WFF have a notably faster k_{fast} rate. The dW2.2 FWF mutant with a tryptophan still close to the heme has a k_{fast} comparable to the WT. Having tryptophan or tyrosine residues near the heme appears to slow the rate of formation of compound I.

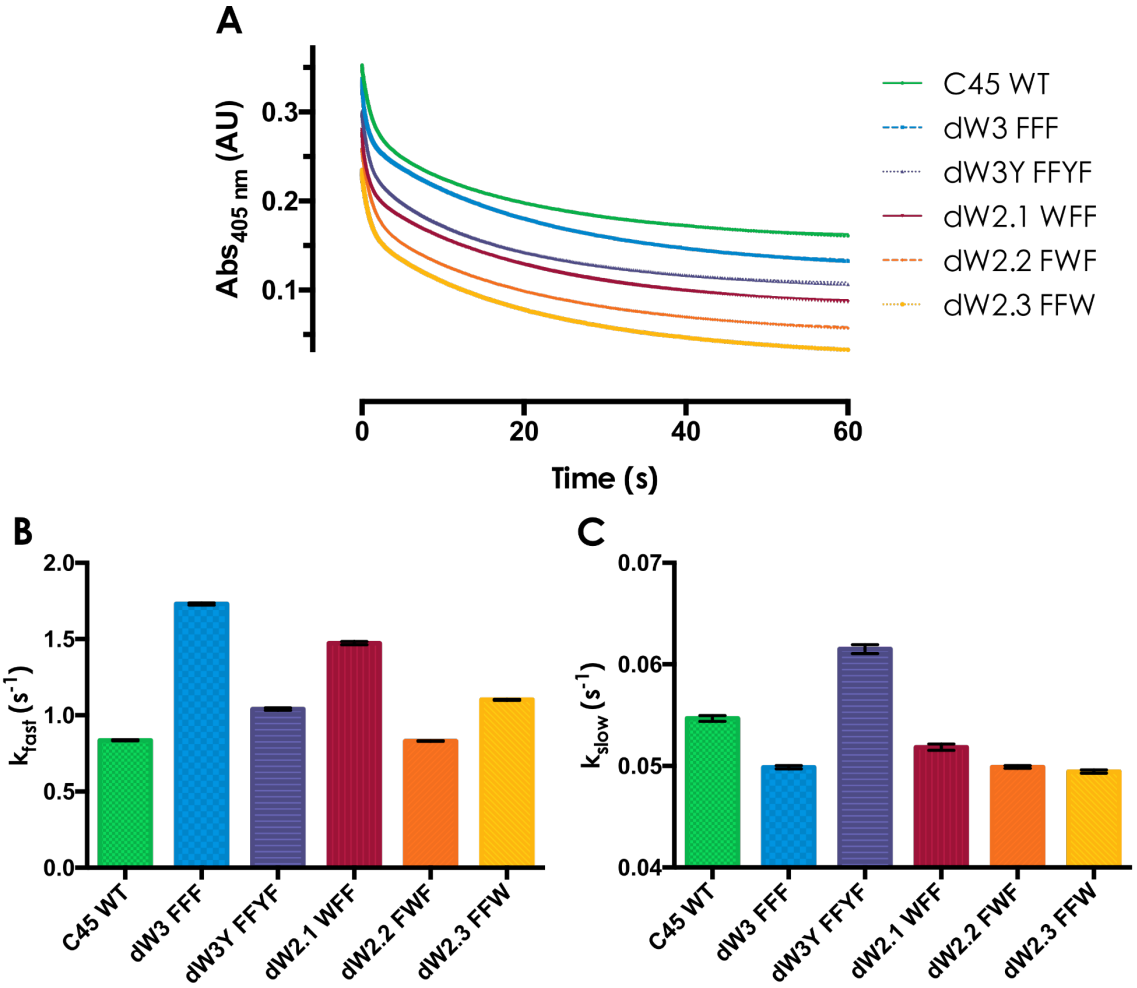


Figure 3.3.13 Soret peak changes on mixing with H_2O_2.
(A) Shows an averaged trace of Soret peak at (405 nm) decrease fitted to a double exponential decay. Traces have been staggered for ease of viewing. (B) and (C) show the k_{fast} and k_{slow} rates from the results of the curve fitting. The error bars show the standard deviation.

	k_{Fast} (s^{-1})		k_{Slow} (s^{-1})	
C45 WT	0.84	± 0.0042	0.055	± 0.00029
dW3 FFF	1.7	± 0.0075	0.05	± 0.00016
dW3Y FFYF	1.0	± 0.0089	0.062	± 0.00044
dW2.1 WFF	1.5	± 0.011	0.052	± 0.0003
dW2.2 FWF	0.83	± 0.002	0.05	± 0.00014
dW2.3 FFW	1.1	± 0.0038	0.049	± 0.00015

Table 3.3.5 Heme degradation rates. *The rates for the fast and slow phases of the decay are tabulated for each protein sample investigated. The ± errors are the standard deviation.*

The k_{slow} rate is increased compared to the WT in the dW3Y mutant with a tyrosine in the distal site above the heme. All other dW mutants have a reduced k_{slow} rate compared to the WT. This suggests that tryptophan residues do not have a protective effect against heme degradation in the absence of an electron donating substrate. The opposite seems more likely

to be the case as the mutant with the closest protein radical-forming residue shows the fastest degradation.

3.3.11 Intermediate formation

Intermediate formation was monitored by the absorbance increase at 520 nm on mixing of 8 μM protein with 800 μM H_2O_2. Figure 3.3.15 shows a comparison of the C45 WT and dW mutants intermediates with HRP, CcP and AXP intermediates. The Soret λ_{max} of HRP is blue shifted on compound I formation and then red shifted in compound II formation [300]. There also is an increase in absorbance in the 600-700 nm range in HRP compound I. HRP compound I forms a porphyrin radical cation whereas CcP forms a stable compound I intermediate with a tryptophan protein radical [301]. The Soret peak of CcP compound I is significantly red shifted and, unlike HRP, it has distinct peaks at 530, 560, and 632 nm [301]. This characteristic signal in the Q-band region of CcP is similar to HRP and AXP compound II signals.

The C45 WT and dW mutants all produce similar intermediate spectra. The Soret λ_{max} in all the maquette intermediates is red shifted compared to the ferric absorbance. The two major peaks in the Q-band region are blue shifted by about 10 nm compared to CcP compound I and HRP and AXP compound II. All the intermediates including the dW3 without any tryptophan residues form an intermediate similar to CcP compound I and HRP and AXP compound II. This would seem to suggest that the isolated maquette intermediate could be compound II rather than compound I. A measure of the definition of the intermediate Q-band peaks is seen in Figure 3.3.15.B. The dW3 mutant has the most defined intermediate spectral peaks with the intermediate peak absorbance (520 nm) greater than the trough (496 nm) by a ratio of 1.21. C45 WT has the least defined spectra followed closely by the mutant with the closest tryptophan (dW2.2 FWF) with ratios of 1.07 and 1.10.

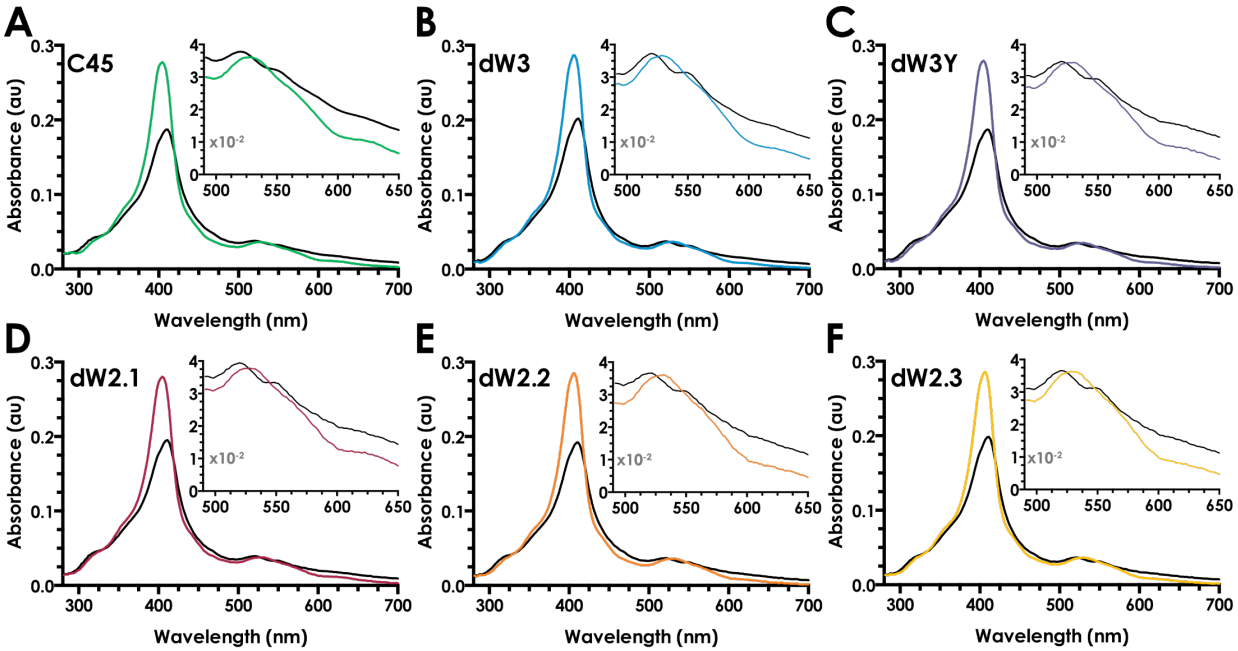


Figure 3.3.14 Intermediate UV-visible spectra. *The coloured trace on each graph shows the ferric resting spectrum at (t = 0.001 seconds). The black trace show the intermediate formed at (t=4 seconds). In each panel an inset enlargement of the Q-band region is included.*

	Peaks (nm)		
	Q1	Q2	Q3
C45 WT	523.0	552.4	572.0
dW3	522.7	550.4	565.7
dW3Y	521.3	551.3	568.5
dW2.1	523.5	551.3	569.0
dW2.2	523.0	550.4	565.0
dW2.3	522.5	550.6	566.5

Table 3.3.6 Q-band intermediate peak positions for dW mutants.

520 nm was seen to have the greatest dynamic range during intermediate formation and was chosen to track formation over time (Figure 3.3.16). Change in 520 nm absorbance, for repeated measurements, were fit to a single-phase exponential association. The results reproduce the trend of the results in Figure 3.3.13.B. The rate of formation of the intermediate is fastest in the dW3 and dW2.1 WFF mutants followed by dW3Y and dW2.3 FFF. The dW2.2 FWF and the WT maquette, which both have a tryptophan at the same end of the bundle as the heme, are the slowest to form the intermediate.

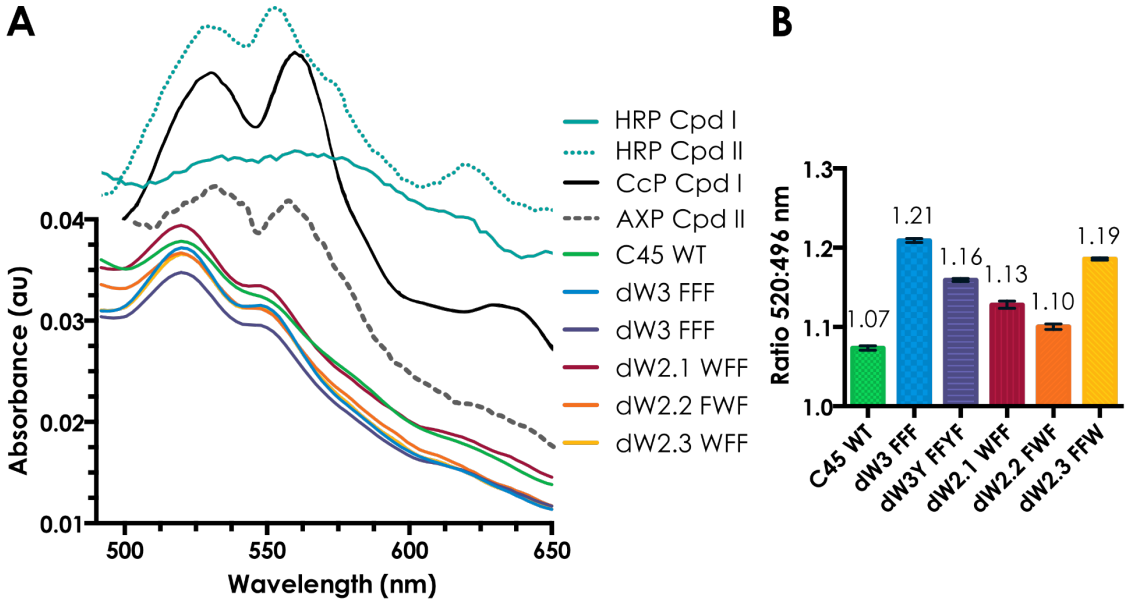


Figure 3.3.15 Intermediate spectra comparison.

(A) *The coloured lines are an overlay of the intermediate at 4 seconds, seen in the inset of the previous figure. The solid and dotted teal lines are HRP compound I (Cpd I) and compound II (Cpd II) respectively, taken from Berglund et al.* [300]. *The black solid and grey dashed lines are the spectra of CcP compound I and AXP compound II respectively, taken from Gumiero et al. at* [301]. *The y-axes of the HRP, CcP and AXP spectra have been scaled to produce this figure. They are not on the same scale as the coloured maquette traces, or each other.* ***(B)*** *shows the ratio of 520 to 496 nm absorbance. These positions are the peak and trough of the upper Q-band region. The ratio gives a measure of the definition of the intermediate heme spectra.*

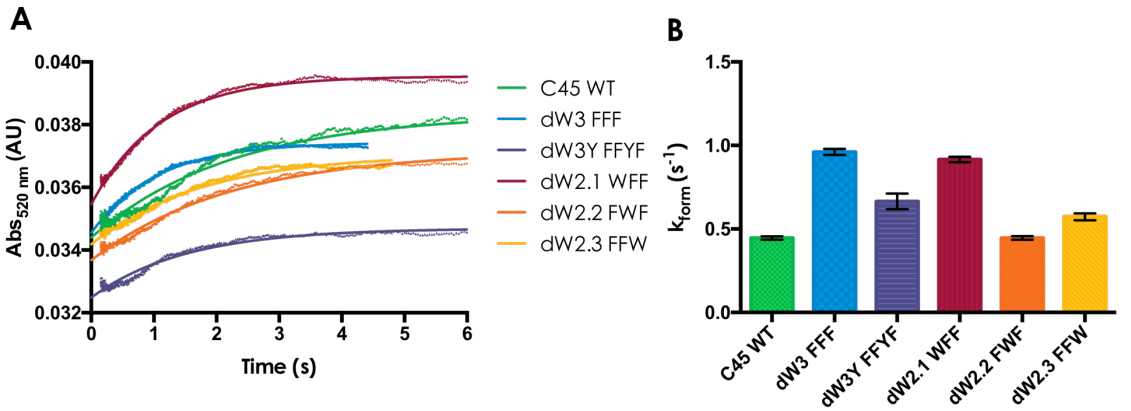


Figure 3.3.16 Rate of formation of reaction intermediate.

(A) *Shows the average spectroscopic changes at 520 nm fit to a single exponential association curve. Only the initial association was plotted, not subsequent decay. Data for the 6 proteins was clipped between 4 and 8 seconds.*

3.3.12 Radical scavenging

Attempts were made to identify a protein radical using a spin scavenging adduct, TEMPO. Myoglobin is known to from a tyrosine radical on mixing with H_2O_2, which can then form a covalent adduct with TEMPO [302]. Myoglobin was used as a positive control for experiments with C45.

The first attempt was made using the Proteomics Facility at Bristol University. Reactions with protein, H_2O_2 and TEMPO were analysed, after a tryptic digest, on nano-LC MSMS. No fragments were detected with the TEMPO adduct on C45 or myoglobin. In both the reaction with H_2O_2 and the controls without H_2O_2, 100% coverage of the C45 and myoglobin sequences was achieved.

A second attempt, without tryptic digestion, was performed using MALDI TOF MS to detect any changes in mass. No mass change was detected between the myoglobin samples with and without H_2O_2 and TEMPO. The same was also true of C45. The C45 mass spectrum after mixing with H_2O_2 and TEMPO had very low signal intensity.

3.3.13 EPR

Electron Paramagnetic Resonance (EPR) spectra were collected for C45 WT and the dW mutants in the resting ferric state and for the radical intermediate. The radical intermediates were generated using peracetic acid (PAA). PAA was used instead of H_2O_2 as it is a stronger oxidising agent and may improve radical signal. Reaction with PAA has been shown to form spectroscopically identical intermediates in all maquettes so far tested [70]. Figure 3.3.18.A shows the EPR features labelled with their g-tensor values. All resting state spectra had the following signals: high spin axial heme iron (s = 5/2 at g = 5.943 and 2.032), low spin heme iron (s = 1/2 at g = 2.265) and high spin non-heme iron (g = 4.270) [286, 303, 304]. The adventitious free iron signal could be a result of iron binding to the negative surface of C45 during the purification. These g-tensor values for these signals were the same (or within the error of the measurement) for all the dW mutants and WT.

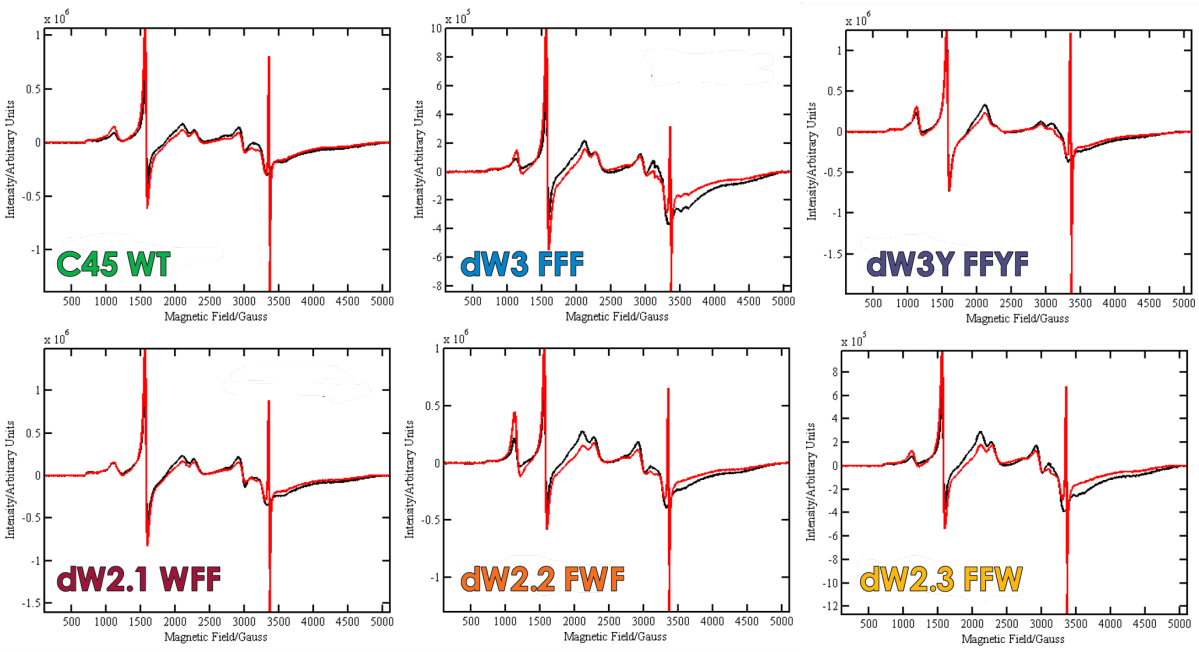


Figure 3.3.17 Wide sweep EPR spectra. *The black line on each graph is the resting state spectrum. The red line shows the spectrum after mixing with PAA and freezing rapidly. Spectra were recorded at 12 K.*

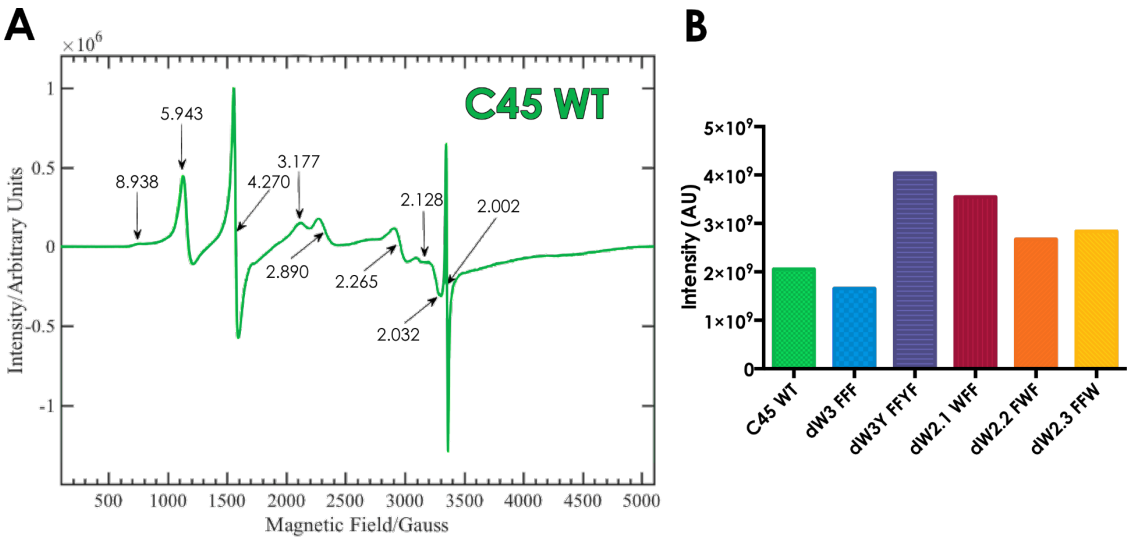


Figure 3.3.18 EPR spectra annotated. ***(A)*** *Is the spectra of C45 WT mixed with PAA. The EPR signals at all magnetic field positions are labelled with their g-tensor value.* ***(B)*** *is a comparison of the amount of radical produced in each sample. It is calculated by taking a double integral of the EPR absorption spectra of each radical.*

All of the dW mutants and the C45 WT produce a radical signal on mixing with PAA, seen in Figure 3.3.18.A labelled with a "g" value of 2.002. WT and all double tryptophan mutants (dW2.1-3) have a "g" value of 2.002 or 2.0019 which is consistent with them all having the same type of radical signal. dW3 and dW3Y radical signals have "g" vales of 2.0017

and 2.0024. This taken with the shape of the radical signal would suggest that they are different radical species. As the same amount of sample was used in each experiment the amount of radical signal can be directly compared. A double integral of the EPR absorption spectra of each radical gives a convenient comparison of the amount of paramagnetic radical signal seen in Figure 3.3.18.B.

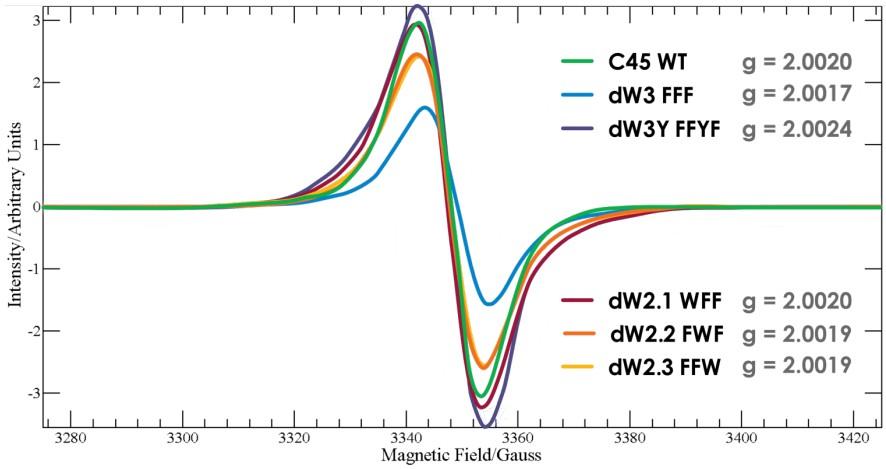


Figure 3.3.19 A comparison of radical signal. *Narrow magnetic field sweeps (150 Gauss) of the radical signal in each sample were taken. The signal was then moddeled (see Appendix C.1) and the g-tensor calculated and displayed for each sample.*

3.4 Discussion

The tryptophan knockout mutants display characteristic UV-visible absorbance and mass changes expected from mutating tryptophan residues to phenylalanine. All mutants are expressed and incorporate heme at a similar level as C45 WT. CD melts shows that all of the proteins are thermostable with highly α-helical structure. The dW3Y mutant has a reduced helical content and a decrease in melting temperature (T_M = 76 ˚C), whereas all other proteins have close to identical amounts of helical content compared to C45 WT (Figure 3.3.5). These results taken with the results for molecular dynamics simulations suggest that the mutants are folded to a similar level to C45.

Generally, MD simulations only sample protein conformations in energetic minima close to the starting structures which may not be the global minima. Without atomistic structural information, it is not possible to say with certainty that structural or topological changes, introduced by making the mutations, are not the cause of any change in behaviour of the mutants. However, the striking similarity in the UV-visible, CD and MD results would lead to the assumption that large changes in the structure and topology of the four-helical bundles has not occurred. If structural changes have occurred, then the CD results seem to suggest the dW3Y had the largest deviation from the C45 WT. The midpoint potential of the dW mutants, other than dW2.1, are within 11 mV of C45. All mutations have caused a more positive shift. These relatively modest shifts are in line with all six maquettes having a *c*-type heme in a similar environment to the WT.

The main trend observed in k_{cat}*, is that having a tryptophan close to the heme has a negative effect (Figure 3.3.11). Having redox-active tryptophan residues far from the heme (WT, dW2.1 WFF and dW2.3 FFW), no tryptophan residues (dW3), and having only an tyrosine above the heme (dW3Y) are all significantly preferable to having a single tryptophan close to the heme (dW2.2 FWF). For all the mutants, there is a decrease in K_M with decreasing the number of tryptophans. For hydrophobic small molecules, the mutations could

cause an increase in binding affinity by creating a more hydrophobic core with slightly less steric bulk.

The result of the k_{cat}* and K_M changes is that mutants without redox-active residues close to the heme show a more than doubling of catalytic efficiency (k_{cat}/K_M) (Figure 3.3.11). This trend has been observed in CcP when the sequential removal of seven redox-active residues resulted in increased catalytic efficiency towards a small molecule substrate [295]. In C45 and its mutants this trend could be explained if substrate binding is not near the protein radical sites or if the proteins radicals are less productive in oxidising electron-donating substrates. If productive substrate binding is predominantly at the heme edge, then formation of the protein radical away from this site could cause a decrease in k_{cat}/K_M. The computational binding study of C45 (in Section 2.3.12) predicted ABTS binding predominately at the heme edge which is in line with the hypothesis that having alternative radical sites may reduce the efficiency of turnover. To test this further it would be interesting to compare steady-state kinetics of TCP, amplex red or other small molecules that are predicted to predominantly bind in other regions of the protein.

To probe the changes to the reaction mechanism further it would be helpful to investigate the H_2O_2 dependant changes and fit the results to a ping-pong mechanism. This might be particularly informative for the dW3Y mutant as the changes to the distal heme surface may well impact on peroxide binding and activation.

The results of the total turnover number experiments would tentatively suggest that the tryptophan residues may make the peroxidase maquettes more resilient to degradation during catalysis (Figure 3.3.12). Tryptophan residues may provide an electron reserve for reaction cycles that occur in the absence of an electron donating substrate. However, these experiments require further repetition and optimisation to be considered reliable. As further work, an assay with glucose oxidase and superoxide dismutase producing a steady supply of H_2O_2 without any other oxygen radicals could be the best way of testing the total turnover number in peroxidase maquettes. Reactions of glucose and glucose oxidase quickly become

limited by the diffusion of oxygen into solution. With the right optimisation, mixing could then be used as a method of controlling H_2O_2 production.

In contrast to the total turnover number experiments, the heme degradation assay showed almost the opposite trend (Figure 3.3.13). When mixing protein and H_2O_2 in the absence of an electron-donating substrate the rate of bleaching of the Soret peak was slowed in all the dW mutants compared to the C45 WT. The exception to this was dW3Y, with a tyrosine above the heme, which degraded quicker than the C45 WT and had the lowest TON. Moving from three to one to no tryptophan residues did not seem to produce an additive effect in the rate of heme degradation. This behaviour would suggest that removing redox-active residues from the protein helps to stabilise the compound I intermediate. This has been observed in CcP, where the rate of compound I decay was seen to decrease as more redox-active residues were removed from the protein [295].

Removing two tryptophan residues from CcP also produced better defined UV-visible absorbance peaks for compound I [295]. This was also seen to be the case with the dW mutants of C45 (Figure 3.3.15). The dW3 mutant had the best-defined compound I absorption peaks in the Q-band region. The dW2.3 FFW, with the tryptophan furthest from the heme, has the second best defined intermediate spectrum. C45 WT and dW2.1 FWF have the worst defined compound I absorption peaks which suggests that tryptophan 43 closest to the heme affects the electronic environment of the heme the most. Other than this change in peak definition, the maximal absorbance of the peaks in all mutants were at very similar wavelengths. The major Q-band absorbance in all proteins under investigation was within 0.5 nm of 523 nm, except for dW3Y which had a slight blue shift to 521.3 nm.

The intermediate species in the *c*-type heme maquettes look most like compound I of CcP or compound II of AXP and HRP (Figure 3.3.15.A). This initially suggested that compound I in the maquette peroxidases may be very short lived and that the spectroscopic observations were of compound II. However, EPR spectroscopy suggested the presence of a radical species

in the maquette intermediate. As compound II does not have a radical species we could speculatively assign the maquette intermediate as compound I.

The EPR spectrum taken of the radical intermediates was not able to conclusively assign which radical is being formed. All the maquettes containing any tryptophan residues produced radical signals with g-tensor values that are the same within the error of the measurement. Somewhat similar spectra have also been obtained for CcP compound I [305]. This could suggest that in a high concentration of oxidising substrate and without an electron-donating substrate even the tryptophan residues 25 and 27 Å away from the heme edge form protein radicals. However, the size of the tryptophan radical signal in CcP at g = 2 is 18-20 G whereas the maquette radical signals at about 8 G [306].

One possible confounding factor, which has not been investigated, is the potential for the maquettes to form dimers. Size exclusion chromatography of the maquettes shows that they do not run as a single species, but denaturing SDS PAGE reveals only one protein. As all the tryptophan residues are at the protein surface even transient association, at the high concentrations necessary for EPR, could lead to radical formation on an associated protein. In the first instance, possible dimerization could be investigated with a computational binding study to give an idea of if dimerization is feasible and the location of possible dimer interfaces. However, C45 has been designed with highly negative surface specifically to discourage dimerization. Additionally, SEC results are not conclusive evidence of dimerisation in solution. The proteins standards for SEC are only a rough estimate of size for proteins that have an extended rather than globular conformation. previous maquettes have shown similar SEC results but do not have a melting temperature that is concentration dependant which would be expected for samples that dimerise [111].

The dW mutants without tryptophan residues have radical signals that have slightly different g-tensor values and shapes (Figure 3.3.19). It was not possible to conclusively assign these signals so the dW3 signal is speculatively assigned as a porphyrin radical. The spectrum of the dW3Y radical only shows an isotropic EPR signal without g-anisotropy or hyperfine

coupling to the beta-methyl protons which are normally observed in tyrosine radical spectra [307]. Dehaloperoxidase is an example of a peroxidase that forms a typical tyrosine radical EPR spectra with hyperfine coupling interactions [308]. dW3Y cannot be assigned as a clear tyrosine radical but its radical signal is different from both the porphyrin radical (dW3) and tryptophan radical (dW2.1-3 and WT) signals.

To conclusively assign the radical species, further work on the radical scavenging experiments is needed. Work to reproduce the radical scavenging detection used for myoglobin and CcP would be the first step. Wright and English were able to detect radical adducts only after tryptic digest using electrospray ionization MS [302]. The preliminary work in this project suggests that using MSMS after digestion is not a suitable method of detecting TEMPO-protein adducts. TEMPO adducts were not seen in ESI MS of intact proteins, which is thought to be a result of the steric hindrance of the protein destabilising the adduct [296]. In this work MALDI TOF MS of haemoglobin and C45 was also not able to detect TEMPO radical adducts, presumably for the same reason.

An alternative hypothesis could be that C45 acts in a similar way to the non-native peroxidase activity of Cytochrome *c* (Cyt *c*). Mitochondrial C*c* is the *c*-type heme protein that is the native substrate for CcP [309]. Under normal conditions the heme in Cyt *c* is bis-ligated by His18 and Met80. During apoptosis there is an increase in Cyt *c* binding and oxidation of an inner-membrane mitochondrial phospholipid called cardiolipin (CL) [310]. The binding of CL to Cyt *c* induces the dissociation of the Met80 ligation on the heme iron increasing its peroxidase activity [311].

Work by Rajagopal *et al*. on a Cyt *c* G41S mutant that displays enhanced peroxidase activity indicates, that the c-type heme is not able to form stable intermediates observed in the classical peroxidase reaction cycles [312]. They were not able to convincingly isolate an intermediate UV-visible spectrum and saw heme degradation in the presence of H_2O_2 and CL [312]. EPR spectroscopy of the peroxidase was not able to assign the g = 2 radical signal to either of the tyrosine residues postulated to be the location of the protein radical. They

concluded that the radical signal is a result of the superimposition of many different protein-bound radicals. They also observed a marked increase in the g = 4.2 free high spin iron signal during peroxidase turnover and rationalised this to be the result of heme degradation [312]. The g = 4.2 signal was not observed in the Cyt *c* control but first appeared after the addition of CL, and then grew further, over 200 seconds after the addition of H_2O_2.

The Cyt *c* and its tyrosine mutants g = 2 radical signals show a very similar trend to what is seen in C45 and its dW mutants. This in combination with the lack of g-anisotropy or hyperfine coupling to the beta-methyl protons, and the failure of the TEMPO radical trapping experiments, could suggest that the radical signal in the maquettes is also a result of the superimposition of many different protein-bound radicals [302, 307, 308]. An arguable difference between C45 and Cyt *c*, is that the g = 4.2 free high spin iron signal in C45 does not show as marked an increase after adding H_2O_2 to resting state enzyme. This could suggest lower levels of heme degradation in the maquettes compared to Cyt *c*.

The picture of radical formation in the maquette peroxidases is a complex one. Like natural peroxidases, C45 and its dW mutants can function with and without tryptophan radicals. One possible explanation of the results of this chapter could be that in the absence of electron-donating substrate all three tryptophan residues on C45 are capable of producing tryptophan radicals. Experiments with an electron donating substrate suggest the more distant tryptophan residues are not as kinetically relevant as the closest tryptophan residue. With ABTS as an electron donor, which is predicted to primarily bind at the heme edge, having redox-active residues producing alternative radical sites reduced the efficiency of turnover. Tryptophan residues in C45 may make the enzyme more resilient (increased TON), though more work is required to develop a reliable assay for this measure.

Chapter 4

Computational investigation of peroxidase protonation state

4.1 Introduction

The *de novo* peroxidase C45 was inspired by the structure of natural peroxidases that have a penta-coordinate heme ligation. This chapter discusses high level QM (Quantum Mechanics) calculations of the reactive cycle of the natural peroxidases. The two peroxidases investigated are Cytochrome *c* Peroxidase (CcP) and Ascorbate Peroxidase (AXP). As discussed in Section 1.8, the role of the distal histidine in the peroxidase reaction mechanism is a subject of ongoing debate. Neutron cryo-crystallography structures of compound I and II, in CcP and AXP respectively, challenged the existing mechanism of proton delivery. In the formation of compound I the distal histidine was assumed to shuttle a proton from the peroxide to form compound 0, and then back to the peroxyferric species to form compound I (Figure 4.1.2.A) [198, 199].

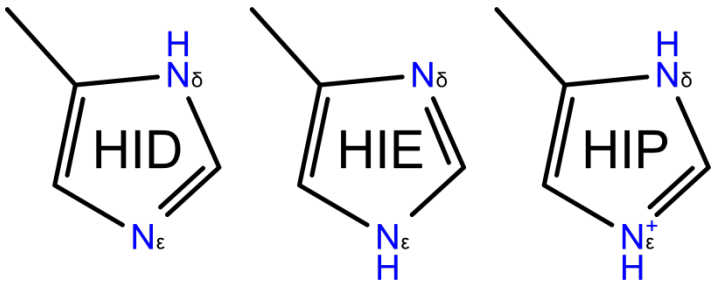


Figure 4.1.1 Naming convention for histidine tautomers in AMBER [229].

The focus of these calculations is to examine the proposed protonation states of the distal site histidine from high resolution neutron cryo-crystallography structures [5, 6]. These structures show that the distal histidine dHIS is in a doubly protonated state (dHIP). To probe the possible mechanistic differences between the two protonation states each species (ferric,

compound I and compound II) was setup in pairs with either dHIP or dHID. Understanding and testing these calculations in the natural systems could then pave the way for their use in models of the *de novo* system, informing potential designed changes in the protein architecture.

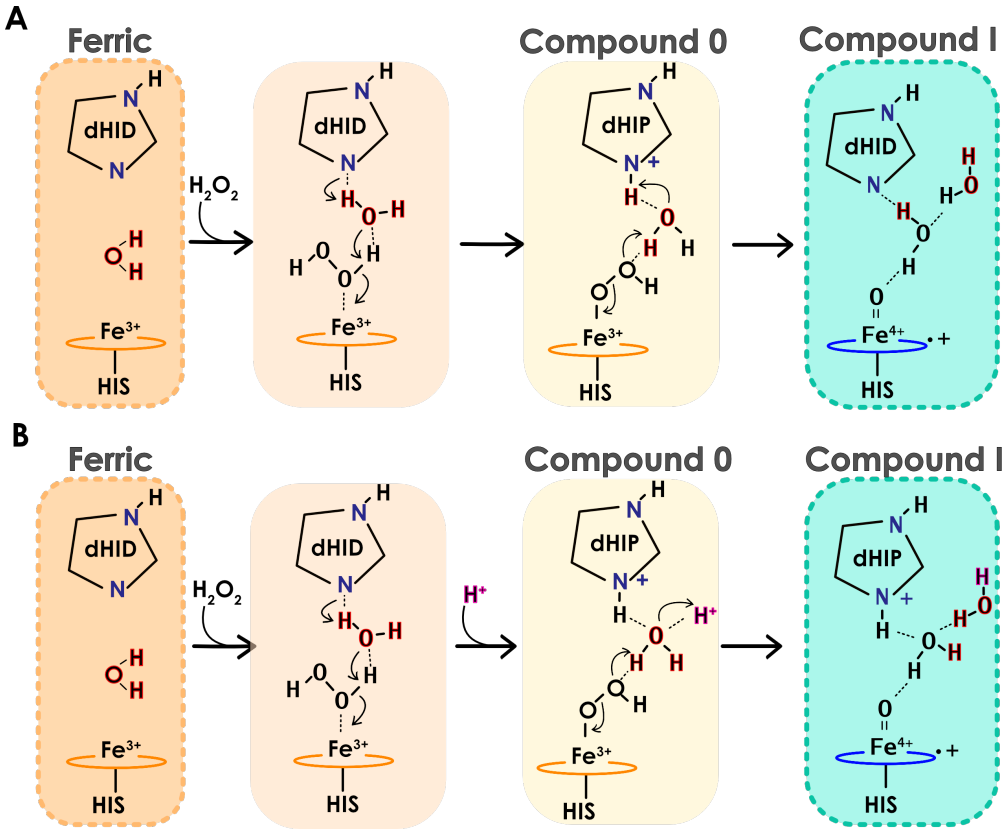


Figure 4.1.2 Compound I formation mechanisms. ***(A)*** *shows the traditional "wet" mechanism* [198, 199]. ***(B)*** *shows the mechanism adjusted for the dHIP protonation state observed in the* neutron *cryo-crystallography structure of compound I [5, 6].*

4.1.1 QM modelling of natural peroxidase systems

To study the subtle changes in the environment of compound I and II intermediates and their reactivity, a description of its electronic structure is needed. Density functional theory (DFT) is a widely used electronic structure method [313]. The advantage of DFT over other approaches such as *ab initio* Hartree-Fock (HF) theory is that the electron density is a function of just three variables, unlike *ab initio* methods where the number of variables describing the wave function is proportional to the number of electrons in the system [314]. The general expression of energy in a DFT calculation is made up of the kinetic energy of the non-interacting electrons, the electron-nucleus attraction energy, exchange-correlation energy and

Coulombic electron-electron repulsion energy. The primary difficulty with DFT is obtaining a suitable expression for the exchange-correlation energy as there is no systematic method of improving the functionals used for the exchange-correlation potential. Some DFT methods, including B3LYP, do not accurately model weak molecular interactions such as Van der Waals forces (dispersion) [315, 316]. Even though DFT calculations are much faster than *ab initio* methods and of comparable accuracy, they are still computationally laborious, limiting the reasonable number of atoms that can be handled by DFT to around 200.

Hybrid density functionals include a mix of Hartree-Fock exchange with DFT exchange-correlation. B3LYP (Becke, three-parameter, Lee-Yang-Parr) is a popular hybrid density functional comprising Becke's 3 parameter exchange correlation functional and Lee–Yang-Parr correlation functional [317, 318]. B3 uses 3 parameters to mix the exact Hartree-Fock exchange correlation and the non-local correlation provided by the LYP expression. The hybrid density functional B3PW91-D3 is also used in this chapter. It uses the same Becke's 3 parameter exchange with the non-local correlation provided by Perdew/Wang 91 and a dispersion correction (D3) component to correct for the known failure of DFT to describe weak molecular interactions [315, 317, 319, 320]. The effect of dispersion and dispersion corrections has also been studied in P450 heme species which are similar to peroxidase heme species [316].

A basis set is a set of functions that represent the molecular orbitals of a system of atoms. The basis set used in this project is the 6-31G* split valence basis with additional polarisation functions to non-hydrogen atoms [321]. For the heme iron, an effective core potential (ECP) was used to overcome the need for larger basis sets to describe the many more electrons of the transition metal. In an ECP the core electrons are replaced with an effective potential. In this project, the heme iron was treated with the LACVP basis set replacing the 1s, 2s and 2p orbitals and with the 3s, 3p, 3d and 4s orbitals treated explicitly [322].

Molecular systems where the valence shell is only partially filled, and the spin of unpaired electrons does not equal one another, are known as open shell systems. These systems have

an excess of either alpha or beta spin electrons, leading to an overall spin not equal to zero. In DFT there are two methods to treat such systems: restricted open-shell and unrestricted open-shell. The restricted treatment uses doubly occupied orbitals as far as possible with the remainder left unpaired. The doubly occupied orbitals are treated identically regardless of whether the unpaired spin is alpha or beta. This fails to consider the effect of spin polarization on the doubly occupied orbitals as unpaired electrons will interact differently with electrons of different or the same spin. Unrestricted open shell methods use two distinct sets of molecular orbitals to describe the alpha and beta spin electrons [323]. Unrestricted open shell methods are more general, as well as being more straightforward to implement computationally ([225] page 108).

4.1.2 QM/MM calculations

The protein environment plays an important role in modulating and stabilising the activity of reactive residues or cofactors in enzymes. Modelling the properties of cofactors containing transition metals can be achieved using high level QM calculations [324]. DFT or *ab initio* methods needed for these calculations are computationally laborious so the number of atoms that can be efficiently modelled is limited. When studying the effect of the protein environment, which is made up of many thousands of atoms, it is necessary to use a hybrid quantum mechanics/molecular mechanics (QM/MM) approach [325]. QM/MM calculations combine the high level electronic calculations on important atoms embedded within a MM calculation of the rest of the protein and solvent environment.

Hybrid QM/MM methods aim to incorporate the benefits of both the QM electronic structures and the rapid molecular mechanics (MM) calculations used for many thousands of atoms [326]. Even a well parameterized MM force field cannot typically model reactivity (making or breaking bonds) and will struggle to describe the more complex electronic environment around transition metals and radicals in conjugated systems. In these cases, QM calculations are needed. For large biomolecules such as enzymes, QM/MM methods offer the capability to embed a QM calculation of a region of interest into the MM environment. This

is typically implemented to include both mechanical and electrostatic interactions across the QM/MM boundary [325, 327]. For biological systems, where the protein environment plays a large role in modulating the activity of the reactive species, having both the spatial and electronic context of the region of interest is clearly very important. The electrostatic and Van der Waals interactions of the MM force field are used to polarize the QM region. The atoms in the QM region are also assigned a MM type Lennard-Jones potential is used to model their steric contribution to the MM region [328].

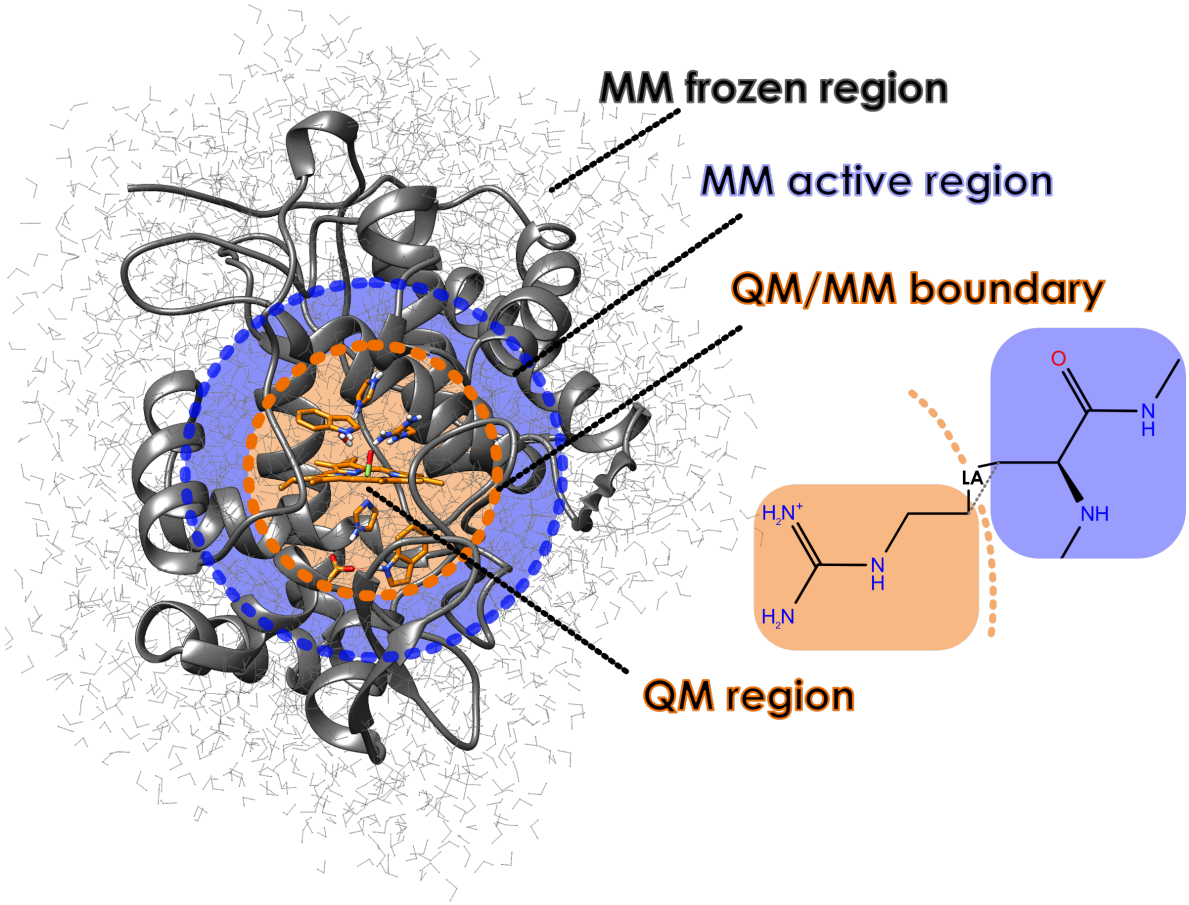


Figure 4.1.3 QM/MM calculation setup. *The atoms treated with a QM calculation are shown in orange. They are embedded in a MM simulation. The two calculations are linked in this case by a "link-atom"method (LA) to satisfy valences over the QM/MM boundary. The MM calculation is split further into and active an frozen region (blue and black). MM active atoms move in response to the QM region (blue) and are surrounded by atoms that have been frozen to set the context for the MM active calculation.*

Semi-empirical QM methods (such as AM1) can be used in QM/MM molecular dynamics simulations typically consisting of many thousands of steps [329]. However, such methods

are unable to model transition metal orbitals. Using DFT in the QM calculation is generally too computationally intensive for dynamic simulations. For heme reactive species discussed here a snapshot of an MD simulation is used in a QM/MM geometry optimization to obtain a total QM/MM energy. This total QM/MM energy is made up of the MM energy, QM energy and QM/MM interaction energy [328].

The boundary between QM and MM regions in enzymes will often fall across a covalent bond. There are a few popular methods for dealing with this issue. This work uses link atoms to satisfy valences over the QM/MM boundary [325, 327]. Here a QM dummy atom, in this case hydrogen, is added to the QM region and some of the non-bonded MM interactions are excluded. The link atom approach could be applied over any covalent atom pair but the best results are achieved on carbon-carbon single bonds away from highly charged groups [328]. In the QM/MM software used in this chapter, ChemShell, charges close to the link atoms are shifted away, and point dipoles added to compensate for the shift [240].

4.1.3 QM cluster models

The first section of this chapter contains QM cluster models of the three-published neutron cryo-crystallography structures [198, 199]. These structures are the ferric (Fe^{3+}) and compound I ($[Fe^{4+}]^{\bullet+}$) species of CcP, and the compound II (Fe^{4+}) intermediate of AXP (PDB id: 4CVI, 4CVJ and 5JPR) [5, 6]. The protonation state of the distal histidine (dHIS) in the neutron structure has dHID in the ferric and dHIP in compound I and II intermediates (see Figure 4.1.1 and Figure 4.1.4).

4.1.4 Setup

Each of the experimental structures formed the starting point for two geometry optimisations: one with the same dHIS protonation as in the neutron structure and a second with the opposite dHIS protonation. Table 4.2.1 shows the 6 structures that were used in the QM cluster models. The doublet and quadruplet spin states of ferric and compound I heme has been seen to be experimentally and computationally degenerate [330]. The choices of

theory level, basis sets, spin states and the size of the QM region to be studied in this work were all informed by previous calculations on the same or similar systems [324, 331, 332]. The protonation states of all ionisable residues were taken directly from the neutron structure of ferric CcP and applied to all equivalent residues in all other setups (PDB id: 4CVI, 4CVJ and 5JPR) [5, 6]. Histidine residues 6 and 175 were HID and 60,96,181 were HIP. Both protonation states of HIS52 were explored.

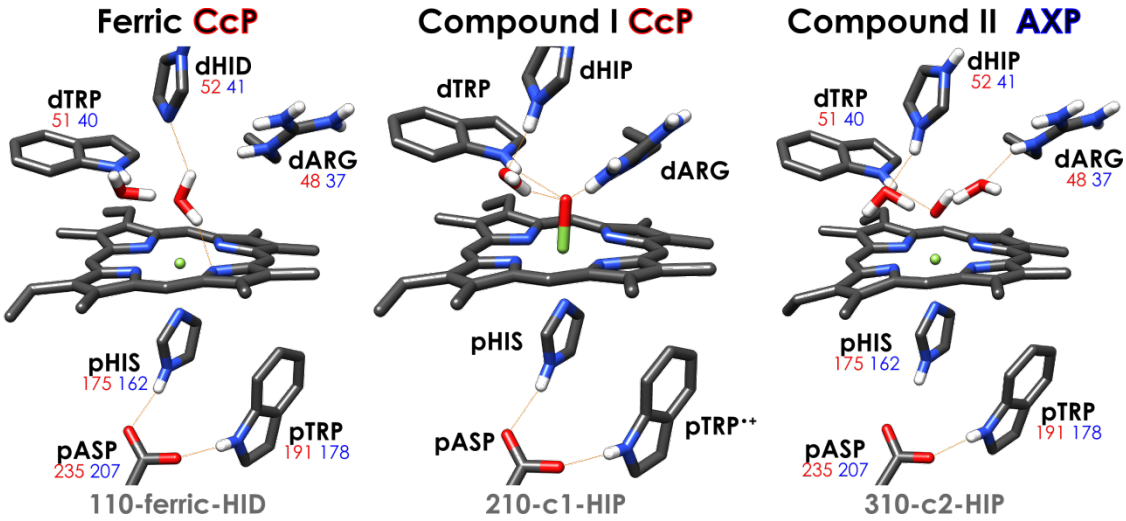


Figure 4.1.4 Neutron structure active site conformation. *The atoms displayed make up the QM cluster and the QM region of the QM/MM calculations (non-polar hydrogens not displayed but included). The coloured numbers under the black residue labels are the residue numbers for CcP (in red) and AXP (in blue). Orange lines represent procedurally detected hydrogen bonds. The grey name below the structures is the file naming index used. The Cγ carbons are fixed for all truncated amino acids except aspartate where the Cβ is fixed.*

4.1.5 An Engineered Ascorbate Peroxidase

The structures investigated with QM/MM were expanded to include the three heme species of each enzyme CcP, AXP and APX2 with and without substrates (see Table 4.2.2). AXP2 is an engineered ascorbate peroxidase that incorporates N-methyl histidine as the proximal heme ligand. It was produced to investigate the role of the proximal hydrogen bonding network, conserved in heme peroxidases. The hydrogen bonding network is between pHIS, pASP and pTRP. The pASP-pHIS interaction in heme peroxidase is thought to give the pHIS partial imidazolate character [333]. The hydrogen bond between pASP-pHIS has been implicated in: fine tuning the redox properties of the heme, aiding O−O bond heterolysis and

stabilizing key oxyferryl intermediates [334–336]. Some QM computational studies have shown that a more imidazolate (more positive) proximal histidine causes a delocalisation of radical spin density to the histidine ligand [337]. However, other QM/MM studies that increased the hydrogen bonding strength of pASP showed little difference in spin density on the pHIS [330].

Experimental results with N-methyl histidine (NMH) ligated AXP (AXP2) have shown that NMH appears to be a more robust ligand than HIS when it comes to peroxidase activity. APX2 with NMH as the proximal ligand increased the turnover number (TON) by a factor of 5 [333]. Combining the kinetic results with the crystallographic structure, Green *et al.* suggest that the NMH is able to improve AXP2 through hyperconjugation of the electron-donating methyl substituent to the Nδ, allowing it to more robustly stabilise radical delocalisation to the proximal ligand [333]. There is particular interest in using N-methyl histidine in the *de novo* peroxidase system as it lacks the natural proximal hydrogen bonding network.

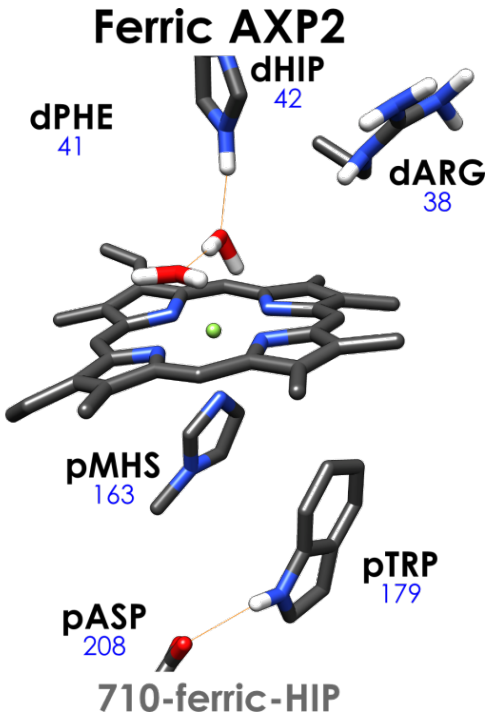


Figure 4.1.5 AXP2 active site conformation. *The atoms displayed make up the QM region of APX2 QM/MM calculations (non-polar hydrogens not displayed but included). The coloured numbers under the black residue labels are the residue numbers. Orange lines represent hydrogen bonds calculated using Chimera. The grey name below the structures is the file naming index used. dPHE label included as a point of reference but was not included in the QM region.*

4.1.6 Generating MM parameters

To run the MM portion of a QM/MM calculation requires all atoms and their interactions to be parameterised in the MM force field. Ferrous and ferric heme B parameters are included as part of the CHARMM force field. Heme B ferric and oxyferryl parameters have been published for cysteine ligated P450-like heme [244]. To rigorously parametrise a new residue or cofactor into a MM force field first requires QM calculations (e.g. in the gas phase) to determine the bond lengths, strengths and the charge distribution of the molecule. The point charges on each atom are generated by a charge fitting method that uses the three-dimensional molecular electrostatic potential, a grid generated by QM calculation [244]. These parameters are then scaled for use within a particular force field. For the CHARMM force field, the charges are scaled to reproduce the quantum mechanically-obtained interaction energies between the solvent and solute [242]. The resulting parameters should then be tested in MD simulations to ensure they accurately model the molecule and its interactions [338].

Full parameterisation of the ferryl hemes was beyond the scope of this project. Instead, updated parameters for histidine ligated ferric, compound I and II heme B were generated using the existing CHARMM27 parameters, QM cluster results and preliminary QM/MM results (Section 4.3.1)[331, 339]. These parameters were only used for short MD simulations and minimisations where the protein movements were constrained before the QM/MM geometry optimisation. In the QM/MM calculations the MM parameters of the atoms in the QM region are ignored, except for their Lennard-Jones potential which is only used by the MM calculation.

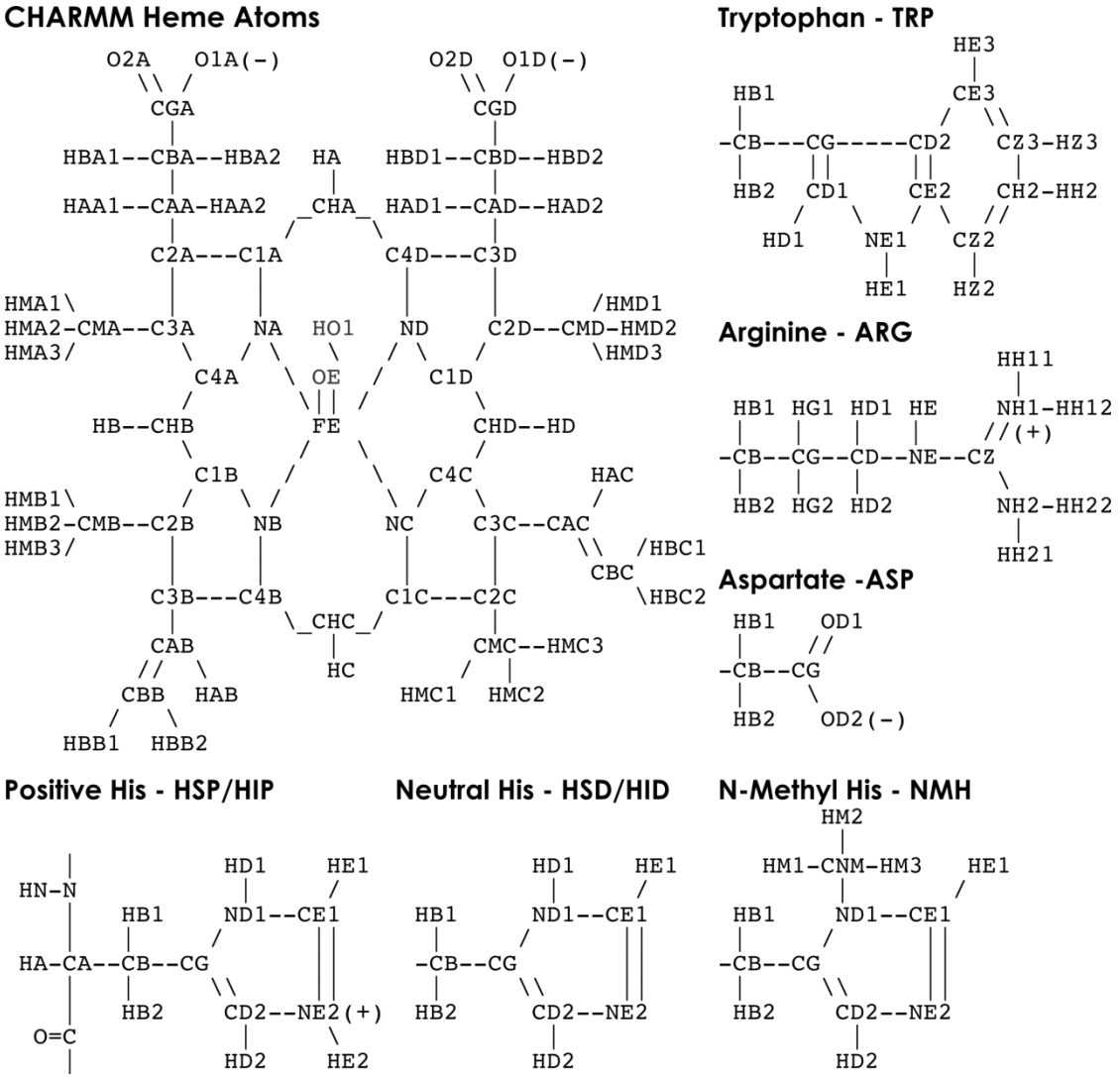


Figure 4.1.6 CHARMM atom naming convention. *The heme atoms include OE which is only present in compound I and II and HO1, only present in compound II. The histidine tautomers are HSP and HSD in the CHARMM naming convention but HIP and HID in the AMBER force fields* [229, 339].

4.2 Materials and methods

4.2.1 Gas-phase quantum mechanics calculations in Jaguar

The high-resolution neutron structures of the ferric, compound I and compound II heme species were used as the starting coordinates (PDB: 4CVI, 4CVJ and 5JPR) [5, 6]. The atoms taken to perform the QM calculations are shown in Figure 4.1.4. High spin iron was chosen for ferric and compound I calculations as it has been seen to be preferable in previous calculations on the same system [324, 331]. Initially, only six Cartesian restraints were used. These fixed the atom nearest to the backbone of each amino acid residue to the position in space it was found in in the experimental structure. This was the Cγ for all residues except aspartate where the Cβ was fixed. Due to convergence issues, an additional dihedral constraint to maintain the porphyrin's rotational orientation about the proximal histidine was added. The Jaguar initial guess was aided by explicitly defining the formal charges of all charged species.

Index	Species	Protein	PDB	Distal HIS	Charge	Multiplicity
110	Ferric	CcP	4CVI	HID	1	4
120	Ferric	CcP	4CVI	HIP^+	2	4
210	Cpd I	CcP	4CVJ	HIP^+	2	4
220	Cpd I	CcP	4CVJ	HID	1	4
310	Cpd II	AxP	5JPR with bound K+	HIP^+	2	3
320	Cpd II	AxP	5JPR with bound K+	HID	1	3

Table 4.2.1 QM cluster calculations structures. *The pdb starting structure, distal histidine residues protonation state, charge and multiplicity are shown. The word compound has been abbreviated to Cpd.*

4.2.1.1 Stepwise QM cluster model setup procedure

1. **Edit protonation**: For 120,220 and 320 the distal histidine (dHIS) protonation state of the neutron structure was edited using UCSF Chimera [340].
2. **AMBER minimisation**: A minimization of the solvent and hydrogens was performed with a strong restraint on all other atoms in AMBER16 using the CHARMM27 force field (maxcyc=600, restraint_wt=100 $kcal.mole^{-1}.Å^{-2}$)[339].
3. **QM Atoms extracted**: The QM atoms were extracted using VMD [237]. Eight hydrogen atoms were added to satisfy the valence remaining from breaking 8 C-C bonds. They were:

two Heme vinyl groups between CAA-CBA and CAD-CBD, one on the proximal aspartate CA-CB and five on the remaining amino acid sidechains between CB-CG.

4. Jaguar 8.5 (Schrödinger, LLC, NY, USA) Setup:
    a. Using Maestro (Schrödinger, LLC, NY, USA) six Cartesian restraints were added to fix the atom that was closest to the backbone for each residue. This was atom CG in all cases apart from ASP where it was CB.
    b. The B3PW91-D3 functional and 6-31G* basis set was used for all atoms other than the heme iron which was described by the LACV3P* basis [315, 317, 319, 320, 322].
    c. Jaguars unrestricted open shell method was implemented with the use of the iufh=1 keyword.
    d. Solvation modelling with a self-consistent reaction field, using the standard Poisson-Boltzmann solver (Isolv=2) was used.

Three different restraint conditions were tested:

The **QM0** setup only used Cartesian restraints mentioned above. **QM1-di** included a dihedral restraint was added between the dHIS and heme to prevent the heme from freely rotating (pHIS:CE1- pHIS:NE2-HEM:Fe-HEM:ND). **QM2-Di-B** included the previous dihedral and a bond restraint on the proximal TRP (NE1-HE1=1). For QM2-Di-B and QM0 the solvation field was applied as on a single point energy calculation after geometry optimisation without it. For QM1-Di the results of geometry optimisation without solvation field were used as a starting point for geometry optimisation with the solvation.

4.2.2 MD and QM/MM

Index	Protein	Species	dHIS	Substrate	PDB	Charge	Mult
120	CcP	Ferric	HIP^+	-	**4CVI**	2	4
110	CcP	Ferric	HID	-	**4CVI**	1	4
160	CcP	Ferric	HIP^+	H_2O_2	**4CVI**	2	4
150	CcP	Ferric	HID	H_2O_2	**4CVI**	1	4
210	CcP	Cpd I	HIP^+	-	**4CVJ**	2	4
220	CcP	Cpd I	HID	-	**4CVJ**	1	4
410	CcP	Cpd II	HIP^+	-	2XJ5	2	3
420	CcP	Cpd II	HID	-	2XJ5	1	3
510	AXP	Ferric	HIP^+	-	1OAF	2	4
520	AXP	Ferric	HID	-	1OAF	1	4
570	AXP	Ferric	HIP^+	H_2O_2 & ASC	1OAF	2	4
580	AXP	Ferric	HID	H_2O_2 & ASC	1OAF	1	4
610	AXP	Cpd I	HIP^+	-	2XI6	2	4
620	AXP	Cpd I	HID	-	2XI6	1	4
630	AXP	Cpd I	HIP^+	ASC	2XI6	2	4
640	AXP	Cpd I	HID	ASC	2XI6	1	4
310	AXP	Cpd II	HIP^+	-	**5JPR**	2	3
320	AXP	Cpd II	HID	-	**5JPR**	1	3
330	AXP	Cpd II	HIP^+	ASC	**5JPR**	2	3
340	AXP	Cpd II	HID	ASC	**5JPR**	1	3
710	AXP2	Ferric	HIP^+	-	5L86	2	4
720	AXP2	Ferric	HID	-	5L86	1	4
770	AXP2	Ferric	HIP^+	H_2O_2 & ASC	5L86	2	4
780	AXP2	Ferric	HID	H_2O_2 & ASC	5L86	1	4
810	AXP2	Cpd I	HIP^+	-	5L86	2	4
820	AXP2	Cpd I	HID	-	5L86	1	4
830	AXP2	Cpd I	HIP^+	ASC	5L86	2	4
840	AXP2	Cpd I	HID	ASC	5L86	1	4
910	AXP2	Cpd II	HIP^+	-	5L86	2	3
920	AXP2	Cpd II	HID	-	5L86	1	3
930	AXP2	Cpd II	HIP^+	ASC	5L86	2	3
940	AXP2	Cpd II	HID	ASC	5L86	1	3

Table 4.2.2 Structures used for MD and QM/MM geometry optimisation.
The PDB codes is bold are the neutron cryo-crystalography structures.

4.2.2.1 Generating parameters

To run the MD simulation and MM portion of the QM/MM geometry optimisation the existing CHARMM heme B parameters needed to be modified. The initial: bond, angle and charge values to add the ferryl oxygen and hydroxide were taken from the QM cluster optimisation results. Lennard-Jones and improper terms were added using similar terms already in the CHARMM27 force field [339]. The bond strengths of the Fe-O were inspired

by AMBER heme B intermediate parameters [244]. The initial values were modified after QM/MM geometry optimisations of the neutron structures. These initial calculations had all the heavy atoms that would make up the QM region fixed (using the AMBER ibelly keyword) during the MM minimisation and solvent equilibration.

Parameters for ascorbate (ASC), hydrogen peroxide (H_2O_2) and the proximal N_δ-methyl histidine (NMH) ligands were also generated. The H_2O_2 parameters were generated using literature values as inspiration but also the results of QM geometry optimisations at the same theory level as the QM/MM calculations [341]. Initial ascorbate parameters were generated using a combination of automatically produced parameters from the CHARMM general force field (CGenFF), using a webserver (SwissParam, www.swissparam.ch), and the AMBER Gaff force field using ACPYPE [247, 342]. These parameters were refined with QM geometry optimisations at the same theory level as used for all calculations in this section. The parameters for 9-methyl guanine and histidine in the CHARMM27 force field were modified to N_δ-methyl histidine [339]. The charges were refined after an initial QM/MM geometry optimisation.

4.2.2.2 Initial calculations

This initial setup lacked proper MM minimisation before QM/MM optimisation and was severely hampered by software problems. The first was a problem in the ChemShell version 3.6 incorrectly reading or implementing AMBER input files. ChemShell then had difficulties interfacing with the ORCA4.0. Finally, transient MPI (Message Passing Interface) errors from an unknown source occurred while sharing nodes with other users on Bluecrystal phase3. The AMBER ibelly function keyword removes the energy of the fixed atoms and therefore is not an ideal option for restraining atoms in MD. Calculations are also slow and cannot be parallelised. For subsequent calculations a Cartesian restraint that can be more easily fine-tuned was used. The AMBER restraint key word “restraint wt” was used along with defining the “restraintmask” atoms.

The list of structures under investigation are the same as those shown in Table 4.2.2. After the modification of the dHIS protonation state was made to the crystal structure, VMD was used to solvate the protein and generate CHARMM parameter (.psf) and coordinate (.pdb) files [237]. AMBERtools16 parmed.py with chAMBER was used to convert these to AMBER compatible input files (parameter.prmtop and coordinate.inpcrd) [229]. The AMBER solvent minimization, solvent heating, system minimization, heating and pressure equilibration were carried out on a local Linux machine using AMBER16 PMEMD for the minimisation steps and PMEMD.cuda for all others. Production MD was carried out with PMEMD.cuda on Bluecrystal phase 3 (University of Bristol HPC, using Nivdia Tesla K20 GPU's).

4.2.2.3 Stepwise QM/MM setup

1. **Edit initial structure**: All atoms were changed to CHARMM atom types. When using the neutron structures, the distal protonation state was either edited or maintained. When using X-ray crystal starting coordinates the protonation for all residues with variable protonation states was set to match the results of the neutron structure. Additional B residues were removed from the pdb file of the neutron structure. Simulations with AXP included potassium bound in the cation binding site. In all other simulations no other ions were added. The X-ray structure of APX2 did not include a cation and none was added. All APX experimental structures were missing two N-terminal residues. The X-ray structure of APX2 was missing three N-terminal residues. In both cases these residues are part of an unstructured loop away from the heme so were not modelled back in. Experimental CcP structures had either one or two N-terminal residues missing; these were modelled back in using Chimera's MODELLER interface (salilab.org/modeller) to make all CcP structures consistent.
2. **Setup and Solvate in VMD**: Using the CHARMM27 parameters and topology files for all amino acids and edited CHARMM27 heme b, Ferric, Compound I (c1), Compound II (c2)

species. VMD generated xplore-type .psf and .pdb files. No ions were added to the setup but ions bound in the starting coordinates were retained.

3. **Convert to AMBER inputs and minimise**: Using ChAMBER in AMBERtools16 parmed.py AMBER .prmtop and .inpcrd files can be generated.
   A standard minimisation protocol for AMBER MD was then performed which included: minimisation of the solvent and hydrogens with strong restraints on heavy atoms, MD heating of solvent with restraints on solute and a minimisation of all atoms.
4. **Equilibration and MD**: MD without any restraints resulted in significant rotamer changes in short MD runs (<1ns). In the region around the heme these were mainly changes in the dHIS rotamer; rotating upwards away from the heme. Different options for restraining the residues around the heme were explored. The following schemes were used for MD equilibration; 50 ps **heating** of system, 200ps **NPT equilibration** and then 10ns NPT **MD production** simulation. Using the AMBER restraints keywords, rather than fixing with ibelly keyword, allows the weight of the restraint to be fine-tuned and the calculation to run GPU parallelized.
5. **Restraints**: A 30 kcal.mol^{-1}.Å^{-2} restraint on all protein residues more that 15 Å from the heme iron was used for all structures except APX2 compound I and II. In the MD there were no restraints on the solvent or residues within 15 Å of the heme iron. APX2 compound I and II MD simulations had no restraints as these structures had the ferric enzyme as a starting structure. For simulations with substrates H_2O_2 and/or ascorbate, weak distance restraints using an AMBER NOE RST file (Nuclear Overhauser Enhancement restraint) file were applied. The restraints are only active if the distance is greater than a cut off value. The restraint energy is calculated as well with a square bottom. Between the distances r2 and r3 the restraint energy is flat, with no restraint applied. The energy between r2 - r1 and r3 - r4 the restraint energy applied increases parabolically. Below r1 and above r4 there is a linear increase in the energy of the restraint. The rk2 and rk3 values are the force constants at the lower and upper bounds (r1 and r4). Setting rk2

= 0 removes the minimum distance restraint. H_2O_2 was weakly restrained between 2.5 and 6 Å from the heme iron and strongly restrained after 6 Å (r1 = 1.30 Å, r2 = 1.80 Å, r3 = 2.50 Å, r4 = 6.00 Å, rk2 = 0 kcal.mol^{-1}.Å^{-2}, rk3 = 5 kcal.mol^{-1}.Å^{-2},). Ascorbate was weakly restrained between 12 and 16 Å from the heme iron and strongly restrained after 16 Å (r1 = 1.10 Å, r2= 1.20 Å, r3 = 12.00 Å, r4 = 16.00 Å, rk2 = 0 kcal.mol^{-1}.Å^{-2}, rk3=5.0 kcal.mol^{-1}.Å^{-2}).

6. **Selecting a representative structure:** A snapshot of the MD simulation was then selected to go forward to QM/MM geometry optimisation. The snapshots were not taken from the first 1 ns of the 10 ns MD. Snapshots with transient movements away from the active site conformation of the experimental structure were avoided. The 10 ns MD trajectories were clustered using AMBERtools16 cpptraj [229]. Five clusters were produced from a fit around the heme and pHIS residue. The clustered structures were used as a guide for choosing the representative snapshot of the MD. The chosen snapshot was converted back to CHARMM atom types and residues as cpptraj only outputs AMBER types.
7. **Minimise:** A MM minimisation of the whole system with the convergence criteria for the energy gradient set to 0.1 kcal.mol^{-1}.Å^{-2} was performed (usually converged around 2000 cycles, max 4000)
8. **Truncate:** VMD was used to select and keep the closest 2000 waters to the protein as a solvent shell. Then the QM, MM active and frozen atoms for the QM/MM calculation were set**.** The **QM region** atoms were the same atoms as in the QM cluster calculation. The heme vinyl groups cut between CAA-CBA, CAD-CBD, Proximal Aspartate cut between CA-CB and remaining sidechains between CB-CG. The two closest water molecules to the heme iron or oxyferryl-O were included in the QM region. The **MM Active region** was defined as same residue as any atom within 4 Å of any QM atoms. To keep the number of MM active atoms consistent, between calculations of the same protein, the active residues for all structures were calculated and compared. A list of protein residues in the MM active

region was then made by amalgamating the calculated residues for each of the proteins together (CcP, AXP and AXP2). The active waters were then also separately defined as the 20 closest water molecules to any of QM region atoms. This allowed all calculations on the same protein to have the same number of MM active region atoms. **Frozen atoms** were defined as all other atoms not in the MM active or QM region.

9. **QM/MM geometry optimisation with ChemShell and ORCA4:** These calculations were performed on Bluecrystal Phase 3 with ORCA4 running the QM calculation parallelised over eight cores. All atoms in the ChemShell geometry optimization were treated with their corresponding CHARMM parameters [240, 343]. The frozen regions atoms are fixed during the geometry optimization but they provide an electronic and steric context for the MM active region. The MM active region atoms run as a MM minimisation using ChemShell's MM engine Dl-poly and the CHARMM force field. These atoms can respond to changes in the QM region of the calculation. For the QM atoms all the MM force field terms, except the Lennard-Jones parameters (which are only used by the MM calculation), are deleted. Hydrogen link atoms are added and an ORCA geometry input is generated with each step. For electrostatic coupling the keyword coupling=shift was used. The ORCA geometry optimization used the B3LYP functional and 6-31G* basis set for all atoms other than the heme iron which was described by the LANL2DZ basis [317, 318, 322]. ORCA keywords convrgence=LooseSCF and dispersion _correction=B3LYP were also used to add dispersion correction (D3) and aid convergence [315].

4.2.2.4 Analysis

Structures and trajectory were analysed and visualised with VMD, using its Tcl scripting interface [237]. Hydrogen bonding calculations and image processing was conducted with Chimera (UCSF Chimera v1.1) [340]. Chimera defines hydrogen bond based on a published survey of crystallographic hydrogen bonds [344]. These strict criteria were relaxed by 20° and 0.2 Å.

4.3 Results

4.3.1 QM Cluster Model

The dihedral restraints added to the QM1 and QM2 setups aimed to keep the heme orientation similar to the experimental structure and stop cycles of the heme rotation. The heme rotating around the proximal ligand resulted in long searches for a converged structure with some simulations taking over 600 cycles. The calculations with the dihedral restraint (QM1-Di) had limited success in this respect as the driving force for sampling different heme rotations proved greater than maintaining the proximal hydrogen bonding network. As a result, the pHIS rotates with the heme producing proximal heme geometries that do not resemble the neutron structure. The calculations with the dihedral constraint did succeed in reaching convergence in fewer cycles.

4.3.2 Iron coordination

The first shell coordination around the heme iron is generally in good agreement with the experimental neutron structure. For the ferric calculations the distance of the Fe-N (HEM:**FE** - pHSD:**NE2**) increase slightly from the neutron structure. The 110-ferric-HID by 0.08 Å and 120-ferric-HIP by 0.13 Å (Table 4.3.2). This could have been driven by a larger movement of the heme which would not be possible in the protein environment. In the compound I and II calculations the heme deviates less from the position in the experimental structures. The Fe-N distances increase by only 0.01 Å in the compound I calculations and less than 0.1 Å in compound II. The Fe=O and Fe-O bond lengths in compound I and II deviate from the experimental structure by less than 0.03 Å. These bond lengths taken from the QM1-Di setup, reproduced with QM0 and QM2 setups, are within the experimental error of the neutron structure. The Fe-O bond lengths support the experimental assignment of a non-protonated compound I with a short Fe=O and a protonated compound II with a longer Fe-OH bond [198, 199].

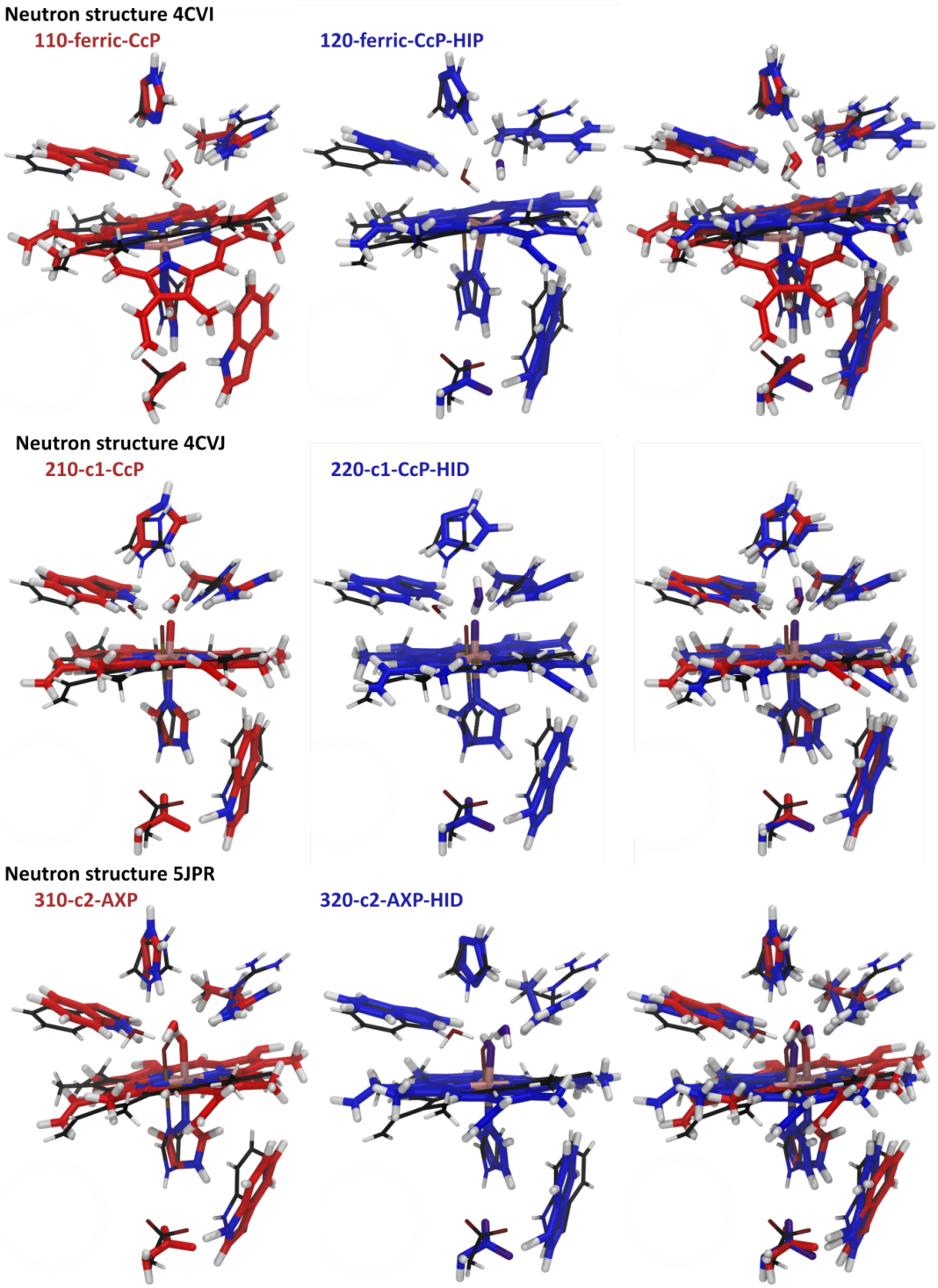


Figure 4.3.1 Structural comparison of results with QM1-Di setup including a solvation field. *Black structures are the experimental structure* [198, 199]. *Red structures are the results of geometry optimisation with the same dHIS protonation as the experimental structure and blue are the results with the alternative protonation state.*

4.3.2.1 Distal Histidine protonation

To compare the QM energy of the change in dHIS protonation of the different species the energy was balanced with a change in protonation of the solvent. It is not possible to

accurately calculate the QM energy of a proton with DFT as it has no electrons. The equations 5.1 to 5.3 show how the energies for each species were balanced. The H_2O and H_3O^+ energies were calculates with the same theory level as the QM cluster models (with and without a solvent field). The grey numbers are the index for that structure.

$$110[Fe^{3+}]HID + H_3O^+ \xrightarrow{\Delta EnF} 120[Fe^{3+}]HIP^+ + H_2O \quad (5.1)$$

$$210[Fe^{4+}]^{\bullet+}HIP^+ + H_2O \xrightarrow{\Delta En1} 220[Fe^{4+}]^{\bullet+}HID + H_3O^+ \quad (5.2)$$

$$310[Fe^{4+}]HIP^+ + H_2O \xrightarrow{\Delta En2} 320[Fe^{4+}]HID + H_3O^+ \quad (5.3)$$

Species	Difference in Energy (kcal/mol)				
	Without solvent field			With solvent field	
	QM0	QM1-Di	QM2-Di-B	QM1-Di	QM2-Di-B
ΔEnF, Ferric, D→P	-41	-35	-11	-72	-75
ΔEn1, Cpd I, P→D	40	49	70	96	125
ΔEn2, Cpd II, P→D	40	47	60	100	111

Table 4.3.1 Difference in energy between QM cluster structures. *The energy differences ΔEnF, ΔEn1 and ΔEn2 correspond to equations 5.1, 5.2 and 5.3.*

The comparison of the total energy of the neutron structure dHIS protonation state ($110[Fe^{3+}]HID$, $210[Fe^{4+}]^{\bullet+}HIP^+$ and $310[Fe^{4+}]HIP^+$) with the alternative dHIS protonation state, suggests that the protonation observed in the experimental structure is energetically favourable for compound I and compound II. However, the ferric structure also seems to favour dHIP which is not seen in the experimental structure.

Table 4.3.1 shows that increasing the number of restraints (QM0 to QM1-Di to QM2-Di-B) resulted in higher energy structures but did not change the trend of the data. Adding a solvation field stabilised further the positive dHIP protonation state but the trend in the results remained the same.

4.3.2.2 Distal Hydrogen bonding network

The dARG in the experimental structures can be seen in two conformations' "in" < 5 Å from the heme iron and "out" > 7 Å from the heme iron (Table 4.3.2). In all ferric QM cluster calculations, the dARG starts in the out conformation and moves in slightly to a position

around 6 Å from the heme iron. The compound I dARG starts 4.3 Å from the heme iron and moves slightly out but remains < 5 Å. The compound II moves out from 7.7 Å much closer and forms a hydrogen bond with the oxygen of the oxyferryl species.

The distal water molecule in the ferric optimisation results has moved away from the heme iron. The experimental structure predicted it to be ligating the heme iron at 2.65 Å (WAT:**O**-HEM:**FE**). In the QM geometry optimisation of the same dHIS protonation as the neutron structure (110-ferric-HID) the water moves out to 3.8 Å from the heme iron (WAT:**O**-HEM:**FE**). In both the dHID and dHIP ferric structures the water has moved to hydrogen bond with both the dHIS and dARG. The first row of Figure 4.3.2 shows the orientation for the ferric heme structures.

The water molecule in the compound I structure has changed orientation more subtly from the neutron structure. In the protonated (210-C1-HIP+) geometry optimisation result the water is slightly closer to the oxyferryl-O forming a hydrogen bond with it as well as the dHIS and dARG. The same is true for the deprotonated result (220-C1-HID) although the water is in a different orientation, acting as a hydrogen bond donor to the dHIS. In the protonated results the distance from the water to the oxyferryl-O and dHIS (HEM:OE - WAT:O and WAT:O - dHIS:NE2) is shorter than the deprotonated (see Table 4.3.2).

The Compound II neutron structure had two water molecules close to the oxyferryl-O with only the closest retained in the QM calculation. In both 310-C2-CcP-HIP and 320-C2-CcP-HID the dARG moves in towards the place of the second water in the neutron structure. The compound II result shows that in both protonation states the water has moved hydrogen closer to the oxyferryl-O forming a hydrogen bond. The protonated (310-C2-HIP) result forms a closer HEM:OE - WAT:O interaction than the deprotonated as well as a hydrogen bond with the dHIS (WAT:O-dHIS:NE2) which is missing in the deprotonated optimised structure.

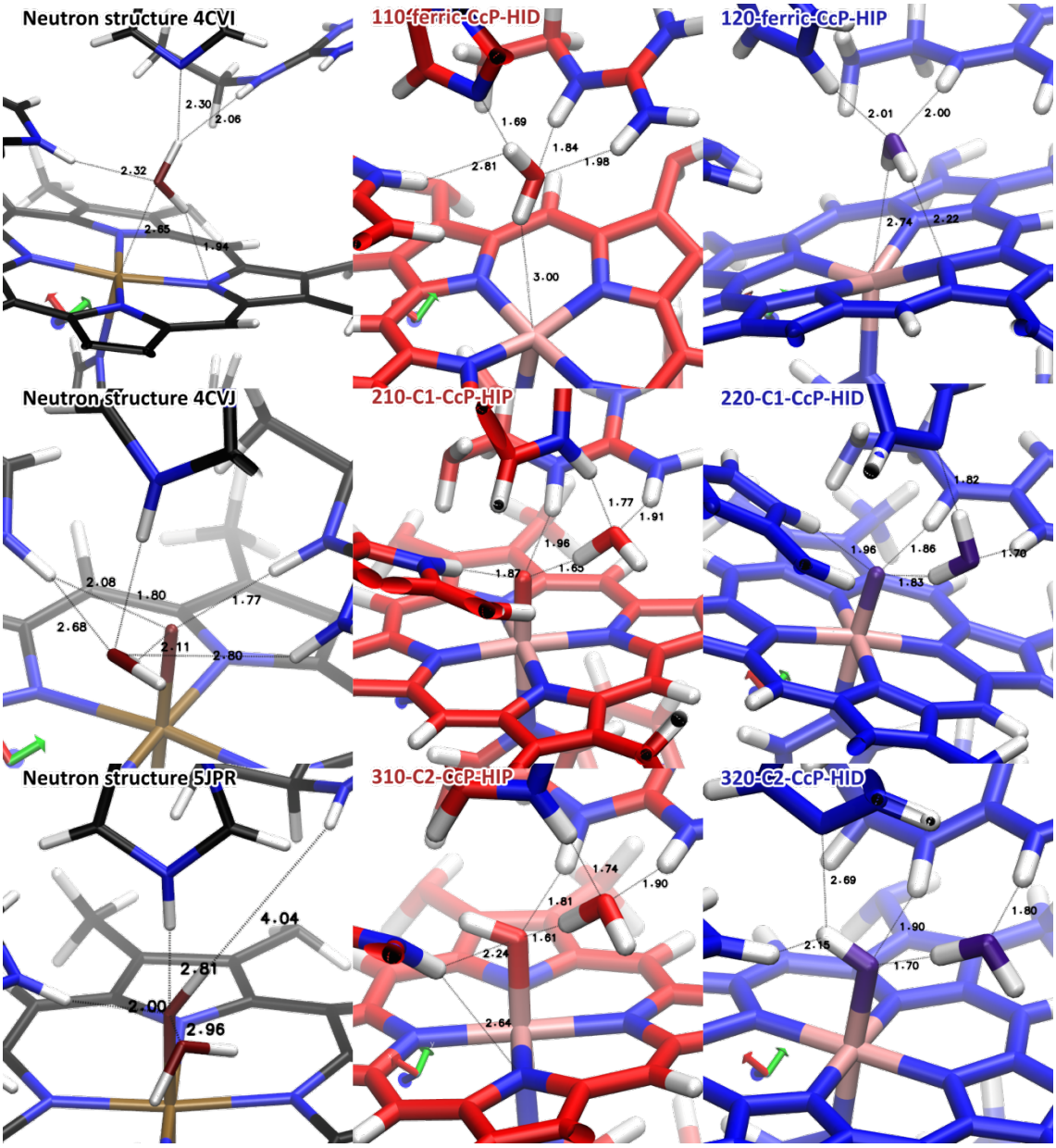


Figure 4.3.2 Distal hydration from QM1-Di minimised structures

Black structures are the experimental structure. Red structures are the results of geometry optimisation with the same dHIS protonation as the experimental structure and blue are the results with the alternative protonation state.

4.3.2.3 Proximal Hydrogen bonding network

The proximal H-bonding network seen in the experimental structure was lost in all of the cluster geometry optimisations. In QM1-Di the rotation of the heme and pHIS was constrained with a dihedral restraint. The energy benefit of heme rotation seemed to outweigh H-bonding pHIS-pASP breaking the interaction and rotating both Heme and pHIS. In the QM1-Di and QM0 the proximal tryptophan hydrogen (pTRP:HE1) transfers to bond with the

ASP OD2 (making ASP-OOH TRP-N+). The exception was *110*-ferric-HID where there was elongation but no transfer. In the QM2-Di-B setup this pTRP N-H bond was fixed as deprotonation does not occur in the protein environment. With the fixed bond the hydrogen bond network is better retained in some of the optimised structures (120-ferric-HID, 210-C1-HIP). The proximal hydrogen bonding network was disrupted in most optimised results. In general the pHIS rotated to face the pHIS:N-H bond directly at the pTRP rather than forming a hydrogen bond with the pASP.

4.3.2.4 Spin Density

The tryptophan radical was observed with high spin density in all optimised compound I structures. One exception to this was in the optimised structure 220-C1-HIP with QM2-Di-B setup where the pHIS had rotated so the pHIS:N-H bond was no longer facing with the pASP or the pTRP. In this structure, the unpaired spin density is found on the oxyferryl species not the pTRP. The deprotonation of the pTRP produced unpaired spin density on the pTRP in all simulations where it occurred.

		Distance (Å)								
Structure		**NS 4CVI**	**110-ferric-CcP-HID**	**120-ferric-CcP-HIP**	**NS 4CVJ**	**210-c1-CcP-HIP**	**220-c1-CcP-HID**	**NS 5JPR**	**310-c2-APX-HIP**	**320-c2-APX-HID**
Cycles		0	89	77	0	79	75	0	86	147
HEM:**FE**	pHSD:**NE2**	**2.02**	**2.10**	**2.15**	**2.12**	**2.13**	**2.13**	**1.97**	**2.04**	**2.06**
HEM:**FE**	dHIS:**NE2**	5.73	6.22	5.78	5.21	5.76	5.53	5.33	5.91	5.02
HEM:**FE**	dTRP:**NE1**	3.72	4.34	3.61	3.15	2.92	3.07	3.26	3.02	3.67
HEM:**FE**	dARG:**NH2**	7.16	5.79	6.29	4.33	5.00	4.81	7.67	4.97	5.34
WAT:**O**	HEM:**FE**	2.65	3.83	3.36	3.92	3.83	3.95	4.15	3.92	4.00
HEM:**FE**	HEM:OE				**1.64**	**1.67**	**1.65**	**1.88**	**1.89**	**1.86**
HEM:**OE**	WAT:**O**				3.06	2.63	2.75	2.96	2.60	2.68
HEM:**OE**	HEM:**HO1**							0.99	0.97	0.96

Table 4.3.2 Distance information from QM1-Di calculations with solvation field. *The Neutron structures (NS) are the first column of each section. The other 6 columns are distances from geometry optimised QM cluster calculations. The shading of data cells is scaled for each row going form the lowest value (red) to highest (blue).*

4.3.3 QM/MM calculations

QM/MM geometry optimisations of the same systems studied in the QM cluster models were setup. This was to observe what difference including the protein environment would make on the optimised structures and to improve the heme and proximal residue movements of the QM cluster model. The number of calculations was also expanded to include each heme species in CcP, AXP and AXP2, as well as calculations including H_2O_2 and ascorbate. Each starting structure was set up in pairs, one with the dHIS protonated (dHIP) and another deprotonated (dHID).

4.3.3.1 Generating parameters

To run MD simulation and MM portion of the QM/MM geometry optimisation the existing CHARMM heme B parameters needed to be modified. The parameters generated for H_2O_2 and ascorbate were tested in short MD (5 ns) simulation with explicit TIP3P solvent and remained in physically realistic conformations. The changes made to the CHARMM heme partial charges are shown in Table 4.3.3. The bond angle between O-Fe-N was seen to be perpendicular to the heme in the compound I and II QM and QM/MM geometry optimisation. However, in the MD simulation of compound II (Fe^{4+}-OH) the O-Fe-N angle would bend seemingly due to electrostatic interactions with the hydroxyferryl-O and H and the porphyrin. To maintain the geometry seen in QM calculation the O-Fe-N angle was strengthened.

Atom Name	Mulliken charges averaged Ferric	Cpd I	Cpd II
HO1			0.37
OE		-0.47	-0.72
FE	0.86	0.67	0.62
NA	-0.7	-0.65	-0.65
NB	-0.7	-0.65	-0.65
NC	-0.7	-0.65	-0.65
ND	-0.7	-0.65	-0.65
C1A	0.23	0.25	0.24
C2A	0.01	0.04	0.03
C3A	0.01	0.04	0.03
C4A	0.23	0.25	0.24
C1B	0.23	0.25	0.24
C2B	0.01	0.04	0.03
C3B	0.01	0.04	0.03
C4B	0.23	0.25	0.24
C1C	0.23	0.25	0.24
C2C	0.01	0.04	0.03
C3C	0.01	0.04	0.03
C4C	0.23	0.25	0.24
C1D	0.23	0.25	0.24
C2D	0.01	0.04	0.03
C3D	0.01	0.04	0.03
C4D	0.23	0.25	0.24
CHA	-0.1	-0.16	-0.18
HA	0.1	0.18	0.15
CHB	-0.1	-0.16	-0.18
HB	0.1	0.18	0.15
CHC	-0.1	-0.16	-0.18
HC	0.1	0.18	0.15
CHD	-0.1	-0.16	-0.18
HD	0.1	0.18	0.15

Table 4.3.3 Charges on central heme atoms. *Mulliken population charges were averaged between symmetrical atoms and two calculations of the same heme species with different dHIS protonation states. See Figure 4.1.6 for atom names. The averaged elementary charge quoted is in Hartree (atomic units). The shading of data is scaled going form the lowest value (red) to highest (blue).*

4.3.3.2 Molecular dynamics

After editing the experimental structures to change the dHIS protonation state, a short MD simulation of the system was run to allow the distal site to sample new solvent conformations. The first round of MD was performed without any restraints. This unrestrained setup showed large movements of the active site residues. For example, in the ferric simulation the dHIS rotamer could change, flipping out of the active sites with the dHIS:Nε around 15 Å from the iron. The pASP and pTRP were seen to move away from the

heme and the pHIS rotated to find new H-bonding pairs. It was assumed that the complex environment around the heme was not being sufficiently captured by the MM force field parameters. The only setup that continued with the unrestrained MD were APX2 compound I and II (index 810 to 940, see Table 4.2.2). This was because the starting coordinates for these simulations was the ferric crystal structure and not a structure of the relevant intermediate. For all other structures (index: 110 to 780) a 30 kcal.mol^{-1}.Å^{-2} restraint on all protein residues more that 15 Å from the heme iron was applied during the MM equilibration and production MD. No restraints were applied to solvent or residues within 15 Å of the heme iron. Simulations with ascorbate and H_2O_2 had distance restraints applied to keep the substrates close to the heme. In the dHIP, simulation a distance restraint applied when H_2O_2 was more than 6 Å from the heme iron was not enough to ensure sampling above the heme iron. The restraint distance was therefore changed to 2.5 Å for all simulations with H_2O_2.

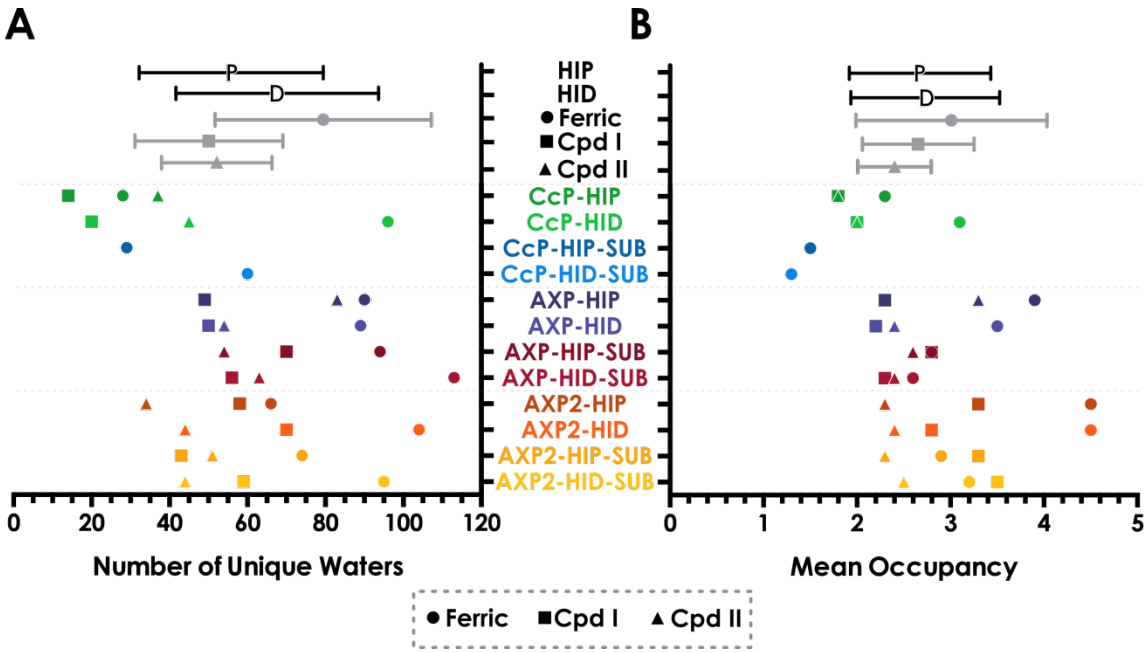


Figure 4.3.3 MD Hydration plots. ***(A)*** *Shows the number of Unique Water molecules that sampled the distal heme face over the course of a 10 ns MD simulations.* ***(B)*** *Is the mean number of water molecules at the distal face of the heme over the course of the MD. The distal heme face was defined as both 6 Å from the heme Iron and 5 Å fron the dHIS. The –SUB results are structure with the substrate H_2O_2 and/or ascorbate, see Table 4.2.2 for details. All points with error bars show the standard deviation and the mean.*

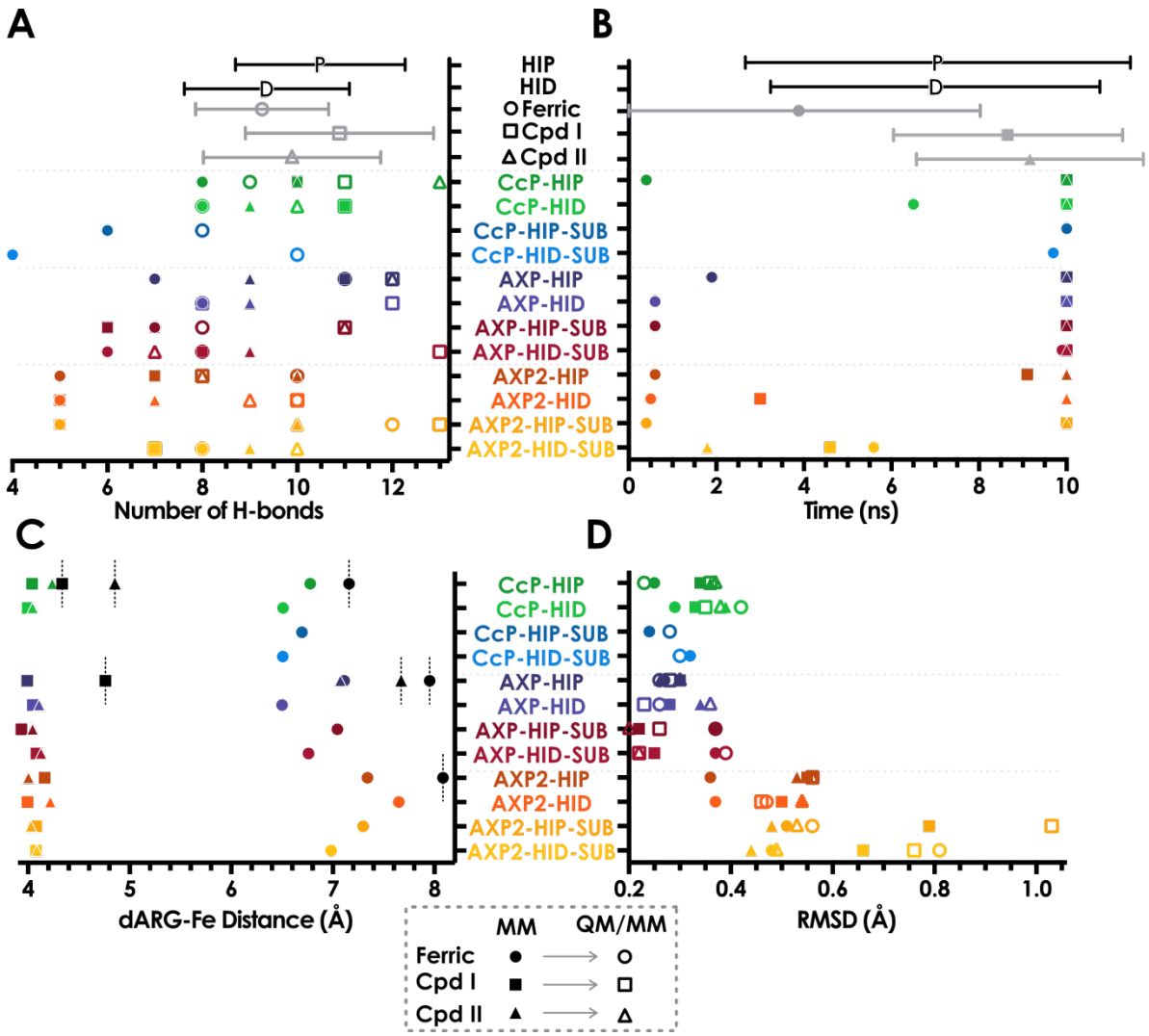


Figure 4.3.4 Hydration and RMSD plots. *A and D are comparison of the MD frame chosen for QM/MM and the result of QM/MM geometry optimisation. (A) is the of the number of hydrogen bonds amongst the atoms of the QM region and water within 8 Å of the iron. (B) Shows the time the closest water to the iron spent as the closest solvent molecule over the 10 ns MD. (C) shows the distance between the dARG:NE2 and HEM:FE for the MD snapshots chosen for QM/MM optimisation. The black points with dashed line are values from the relevant experimental structure. (D) Is the RMSD of the heavy atoms of the QM region. Both the MM structure and QM/MM results are compared to the experimental starting structure. The –SUB results are structure with the substrate H_2O_2 and/or ascorbate. See Table 4.2.2 for details. All points with error bars show the mean and the standard deviation.*

Figure 4.3.3 shows that on average, MD simulations with ferric heme have more water molecules accessing the distal site and that it is also occupied by more water on average (Figure 4.3.3.B). There is no significant difference between the average number of water molecules occupying the distal site of dHIP and dHID setups. The variance in the hydration results decreases going from high variation in the ferric systems to lower in compound I to II

systems (Figure 4.3.3). In the majority of compound I and II MD simulations the closest water remains the closest water to the heme iron for the entire simulation (Figure 4.3.4.B). These trends in hydration suggest that the solvation in the Oxy- and Hydroxy-ferryl intermediates is more ordered or more constrained. Most of the RMSD of the final QM/MM result compared to the experimental structure is introduced during the MD simulation. Figure 4.3.4.D shows the majority of the RMSD increase from the experimental structure occur during the MD simulation, with only small changes occurring during QM/MM optimisation (on average a 8.2% change, sd= 19.6).

A prominent portion of the RMSD of the QM region atoms was the movement of the dARG. During the 10 ns MD of the ferric simulations (index; 100's, 500's and 700's) the "in" conformation of the dARG was sampled on average 13% of the time (sd=23). The "in" conformation was sampled 82% and 85% of the time in the compound I and II MD simulations (sd=21 and 28). The "in" conformation was defined as when the dARG:NE2 was closer than 5 Å to the heme iron. For simulations of the ferric enzyme dARG started in the "out" position and on only 2 occasions moved in. The compound I simulation started and ended with the dARG "in" conformation, except in the AXP2 simulation where the dARG started out and moved in. All compound II simulations where the dARG started "out" it moved "in" over the course of the MD.

4.3.3.3 Convergence

Many of the ferric QM/MM geometry optimisations took a large number of cycles to converge (>400 cycles). For 3 Ferric minimisations (110, 150 and 710) the criteria had to be loosened after 1000 cycles. The second row of Figure 4.3.8 shows that some QM/MM results from alternative MD snapshot structures suffered convergence issues. Part of the issue seems to be that the hydration of the ferric heme is more varied and selecting the most appropriate waters for the QM region becomes more difficult.

4.3.3.4 Iron coordination

The Fe=O bond in compound I is reliably reproduced in the QM/MM geometry optimised structures (see Figure 4.3.5.A). All the CcP and AXP results have an average Fe=O bond length 1.680 Å (sd=0.0025). The AXP2 bond is slightly but significantly shorter (1.665 Å, sd=0.0017, Pvalue= 0.0095). All these values are slightly longer (~0.3 Å) than the experimental structure and within the same range as the QM cluster models. The compound II Fe-OH bond distances are also reliably reproduced by the QM/MM result. The data follow a trend of decreasing Fe-OH bond length ongoing from CcP to AXP to AXP2.

The Fe-pHIS:NE2 (Fe-N) bond lengths in the QM/MM optimised results were more variable than F-O bond lengths but still broadly in line with the experimental results.

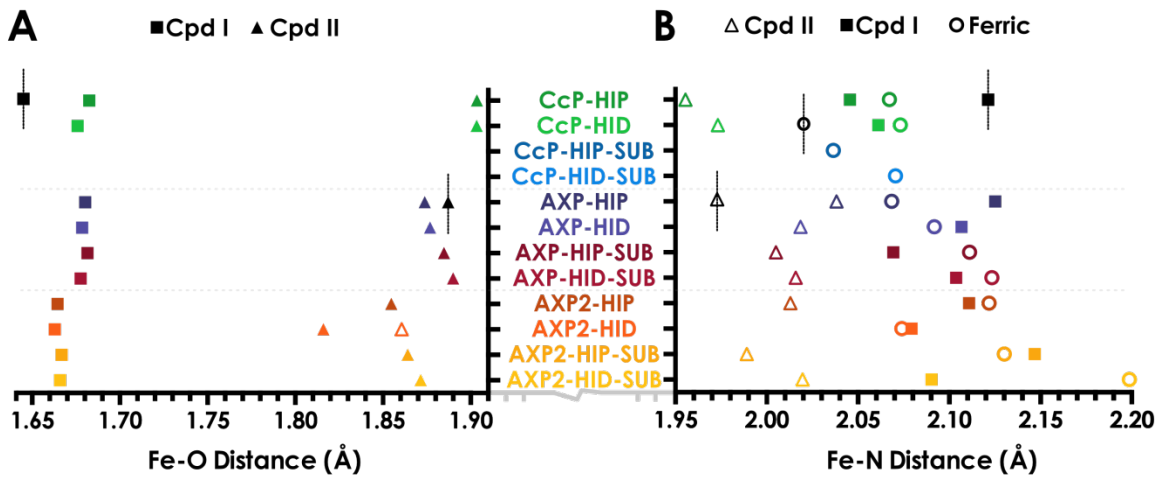


Figure 4.3.5 Iron oxygen and iron nitrogen bond distances. ***(A)*** *shows the distance between the heme iron and oxygen in the ferryl intermediates. One value for AXP2-HID is from graph* ***B****.* ***(B)*** *shows the distance between the HEM:Fe and pHIS:NE2. All coloured plots are the results the QM/MM calculations. The black points with dashed lines are values from the neutron cryo-crystallography experimental structures.*

4.3.3.5 Distal histidine energy changes

The same procedure of balancing final energies conducted with the QM cluster models was also conducted for the QM/MM energy of optimised structures. For the QM/MM results, equation 5.1 has been reversed, producing equation 5.4. The energy difference quoted for the QM/MM results is always comparing moving from dHIP to dHID. This is unlike the QM cluster results where the energy of the neutron structure protonation state was compared

with the modified protonation state. In most cases, the difference in energy of the QM/MM results is much greater than expected for a simple proton affinity, by an order of magnitude in some cases. On average the strongest preference for dHIP is in the compound II structures and the lowest energy barrier in the ferric calculations. The only calculation where the pHID is energetically favourable is with CcP in the ferric form with H_2O_2 as a substrate.

Alternative snapshots in the second row of Figure 4.3.8 show minimisation of systems that did not converge after >1000 cycles. These structures are all dHID starting structures were the distal site water or H_2O_2 has deprotonated over the course of the minimisation. The energy differences between dHIP to dHID are relatively low (shown in Figure 4.3.8, not Figure 4.3.6) compared to the average difference of the QM/MM calculations.

$$[Fe^{3+}]HIP^{+} + H_2O^{+} \xrightarrow{\Delta EnF} [Fe^{3+}]HID + H_3O \qquad (5.4)$$

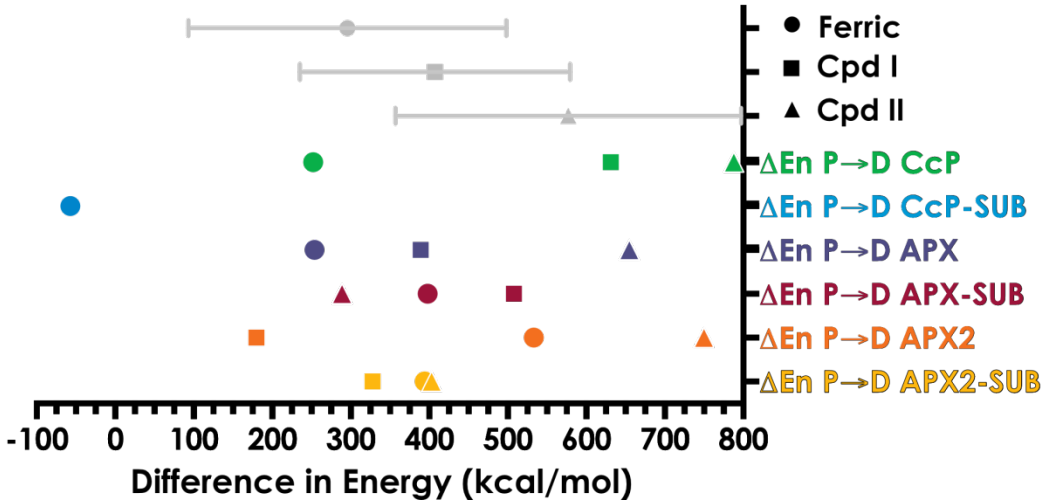


Figure 4.3.6 QM/MM energy differences to move from dHIP to dHID. *All points with error bars show the mean and the standard deviation.*

4.3.3.6 Distal Hydrogen bonding

Significantly more hydrogen bonds were detected in the QM/MM results compared to the MD input snapshot. The average detected hydrogen bonds from MD input snapshot was 7.63 and QM/MM results average was 9.88 (sd = 1.91 and 1.83 Pvalue = <0.005). Figure 4.3.4.A shows that compound I QM/MM results have on average the most hydrogen bonds, then compound II, finally ferric structures. dARG can be seen to form at least one hydrogen bond with compound I oxyferryl-O in all QM/MM optimised structures. This is also the case in over half the compound II structures. Figure 4.3.7.A shows that for all compound I and most of the compound II QM/MM structures the dARD is in the "in" partition. The QM/MM

Fe-dARG distances are all slightly longer (~0.5) than the MD input structure and are more in line with the experimental results.

The distal dHIS-Fe distances are very similar for all QM/MM results. The nature of the hydrogen bond that dHIS makes with solvent in the distal site varies depending on the protonation state it is in. The closest water to the ferryl-O in Compound I and II is highly ordered in the QM/MM optimised results. In all dHIP calculations, this distal water donates a hydrogen bond to the ferryl-O and accepts a hydrogen bond from the dHIP. In all dHID calculations the water donates a hydrogen bond to both the ferryl-O and the dHID.

The hydrogen bonding of dTRP to the ferryl-O of both compound I and II is observed in all CcP and AXP calculations. The dTRP-Fe distance is shorter on average in the compound I and II calculations than the ferric. The compound I calculations show very little variation from each other and the experimental neutron structure (Figure 4.3.7.C).

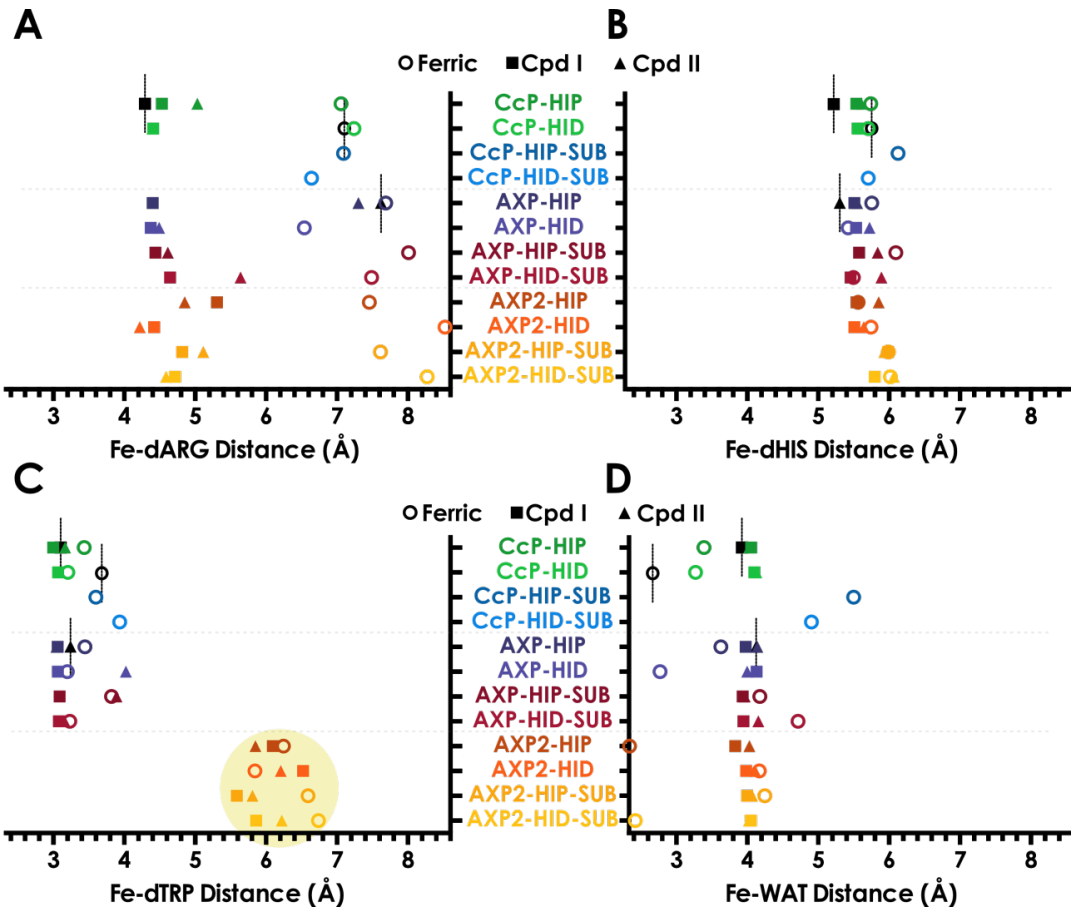


Figure 4.3.7 QM/MM result distance measures. ***(A)*** *shows the distance between HEM:Fe and dARG:NH2.* ***(B)*** *shows the distance between HEM:Fe and dHIS:NE2.* ***(C)*** *shows the distance between HEM:Fe and dTRP:NE1. APX2 has a dPHE rather than dTRP. The results in the yellow shaded circle are the distance to the dPHE.* ***(D)*** *shows the distance between HEM:Fe and WAT:O. All coloured points are the results the QM/MM calculations. The black points with dashed lines are values from the neutron cryo-crystallography experimental structures.*

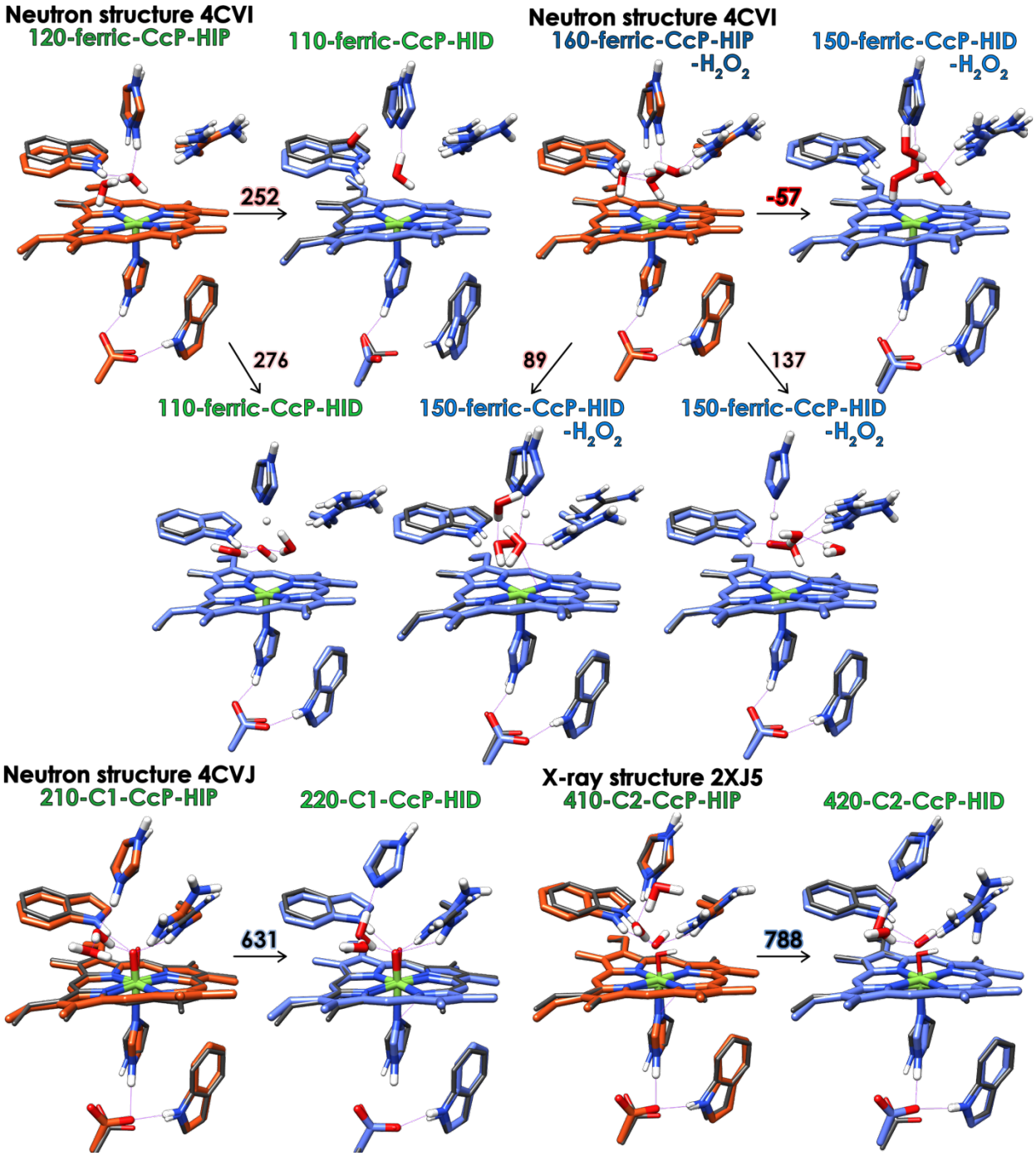


Figure 4.3.8 CcP QM/MM optimised structures. *Only the heavy atoms (and polar hydrogens) of the QM region are displayed. QM/MM geometry optimisation results are in colour and are displayed with experimental starting structures in black. Structures in the second row have not converged. dHIP results are in Orange and dHID are in blue. The thin purple lines are the calculated hydrogen bonding interactions. The numbers above arrows, going between dHIP to dHID pairs, are the balanced QM/MM energy differences (kcal/mol). The structure titles are coloured to match the graphs in this section.*

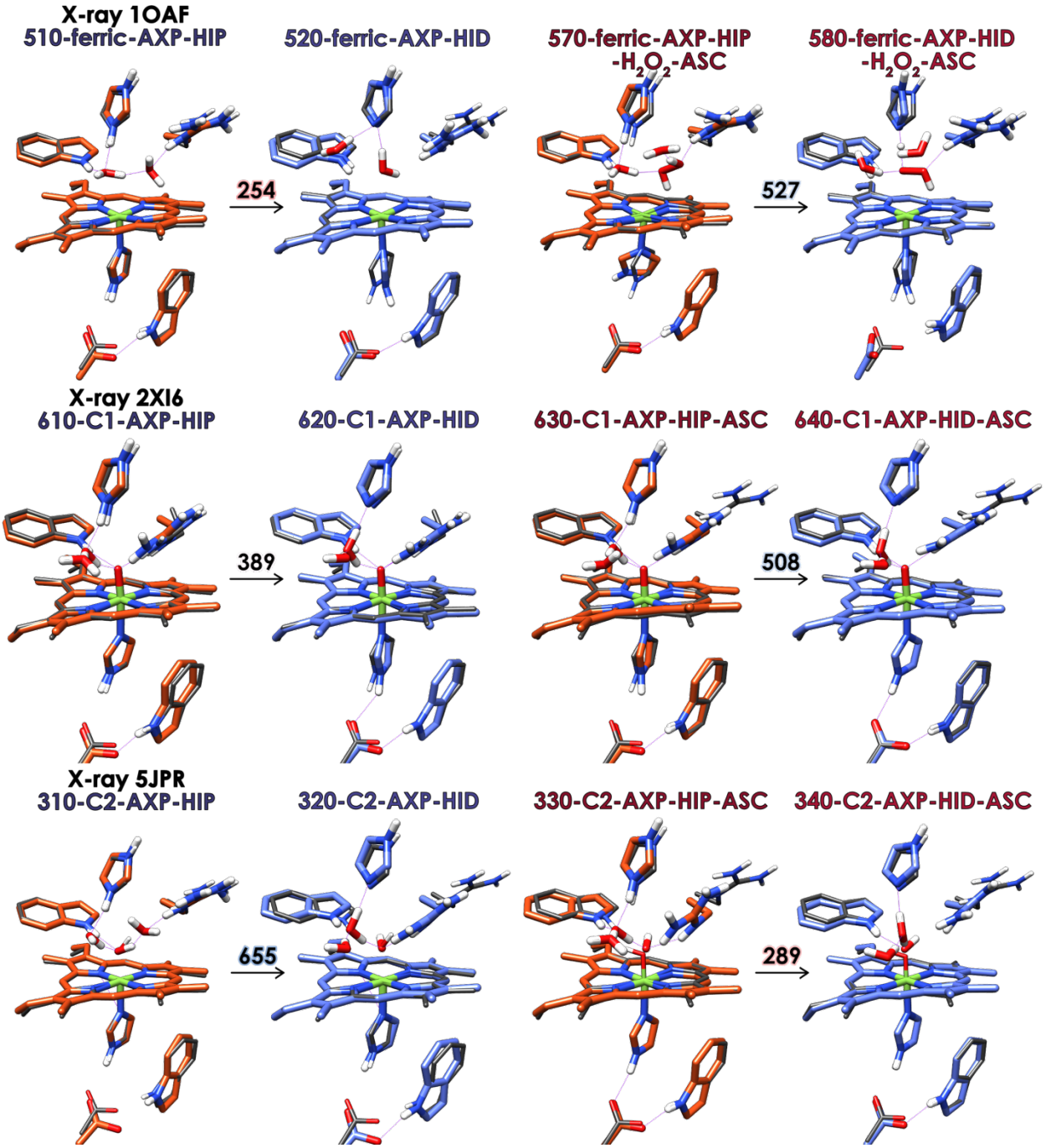


Figure 4.3.9 AXP QM/MM optimised structures. *Only the heavy atoms (and polar hydrogens) of the QM region are displayed. QM/MM geometry optimisation results are in colour and are displayed with experimental starting structures in black. dHIP results are in Orange and dHID are in blue. The thin purple lines are the calculated hydrogen bonding interactions. The numbers above arrows, going between dHIP to dHID pairs, are the balanced QM/MM energy differences (kcal/mol). The structure titles are coloured to match the graphs in this section.*

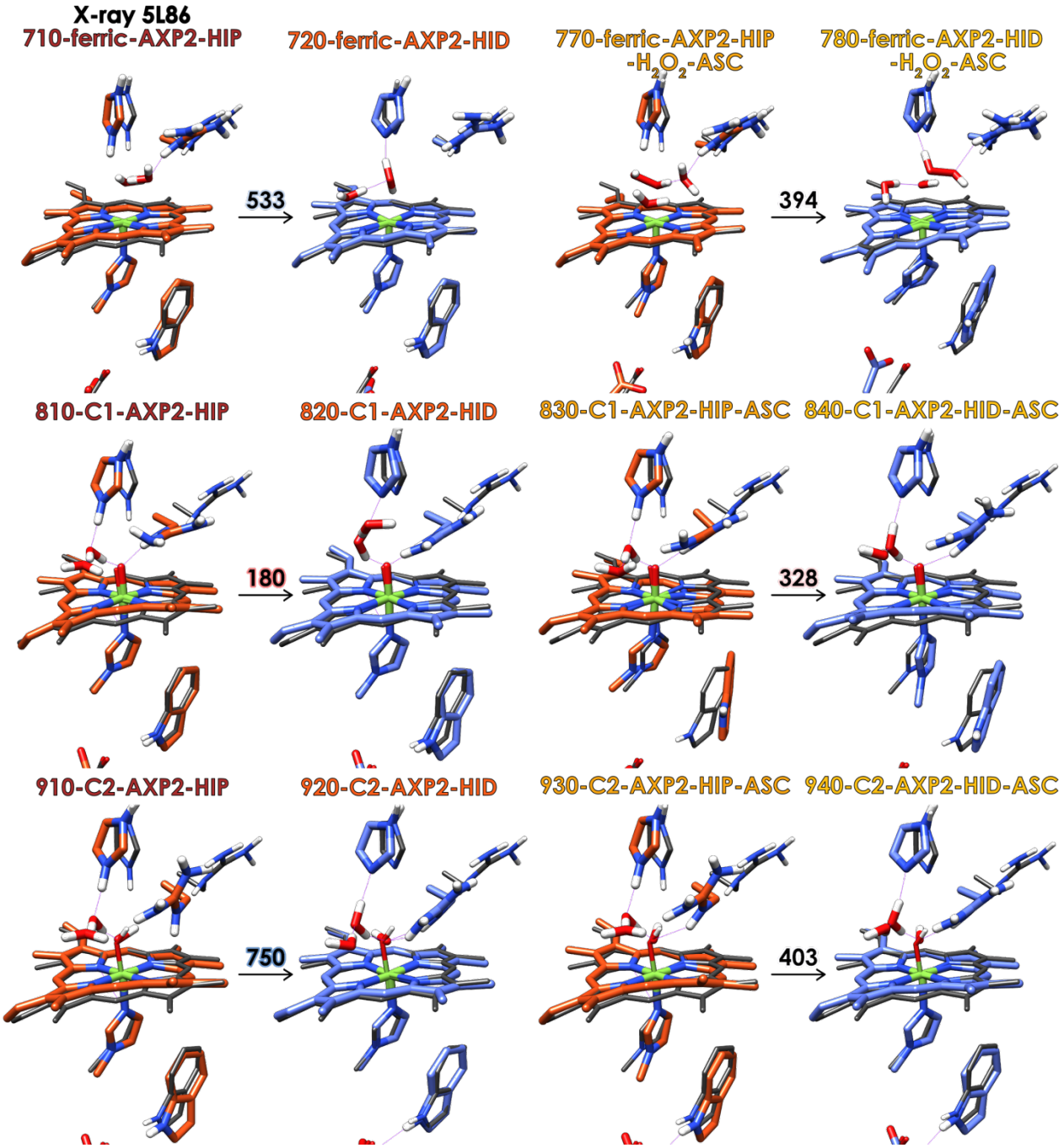


Figure 4.3.10 AXP2 QM/MM optimised structures. *Only the heavy atoms (and polar hydrogens) of the QM region are displayed. QM/MM geometry optimisation results are in colour and are displayed with experimental starting structures in black. dHIP results are in Orange and dHID are in blue. The thin purple lines are the calculated hydrogen bonding interactions. The numbers above arrows, going between dHIP to dHID pairs, are the balanced QM/MM energy differences (kcal/mol). The structure titles are coloured to match the graphs in this section.*

4.3.3.7 Proximal hydrogen Bonding

The proximal hydrogen bonding network is generally well maintained by the QM/MM calculation compared to the neutron structure. The pASP-pTRP hydrogen bond is detected in

all but one of the CcP results and all but two of the AXP results. Rotation of the pHIS in the QM cluster calculation is seen to a lesser extent in the QM/MM calculations. In all CcP compound I and II calculations, the pHIS rotates to donate a hydrogen bond on to the same pASP-O that is accepting the pTRP hydrogen bond. This behaviour is not seen in the AXP and AXP2 calculation. Where movements of the pHIS in AXP and AXP2 calculations occur they have been introduced during the MD equilibration before the QM/MM optimisation. In the QM/MM results where the pHIS-pASP hydrogen bond is not detected, the pASP is forming a hydrogen bond with a water molecule. The water molecule in this position near the pASP and pHIS is seen in the experimental structures and QM/MM results.

4.3.3.8 Spin Density

No difference was seen between the spin density on pHIS of CcP and AXP and the pNMH of AXP2 in compound I calculations. In all ferric calculations, the pHID or pNMH has a small amount of spin density (0.1 au). In compound I and II calculations the only notable spin on the proximal ligand is very small negative spin (- 0.01 au) on the iron ligating nitrogen, NE2. Figure 4.3.12.B-D shows that in AXP2 there is a significant reduction of spin density on the compound I iron and an increase on the oxyferryl-O.

Figure 4.3.11 shows that the QM/MM experiments can reproduce the experimental result of a pTRP radical seen in CcP compound I. The radical compound I signal on the pTRP in AXP is about half that seen in CcP and in AXP2 the spin density is reduced further.

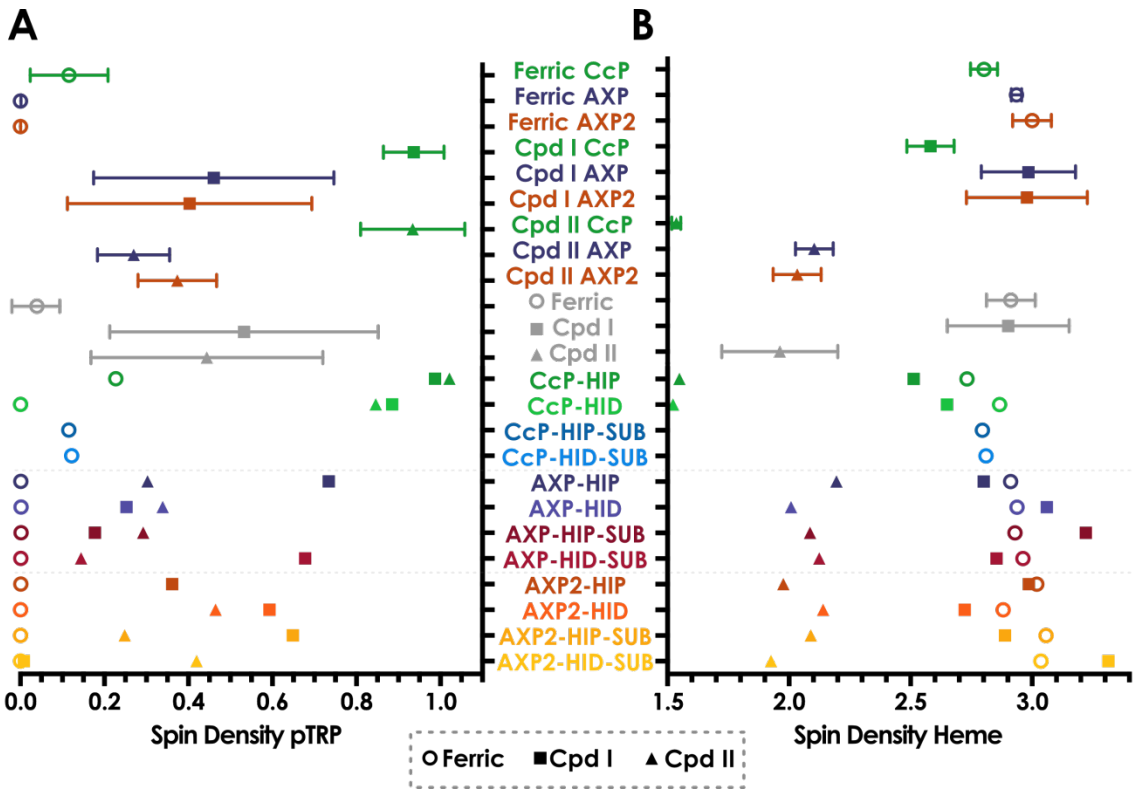


Figure 4.3.11 Spin density on heme and pTRP. *The plotted vales are the sum of the magnitude of spin density on the pTRP* ***(A)*** *and heme* ***(B)****, it is therefore possible for the value to be greater that the total spin of the system. All points with error bars show the mean and the standard deviation.*

A Mulliken population analysis was performed to calculate atomic charges for the QM region atoms. The iron charges in the compound I and II intermediates are more negative when compared to AXP and CcP (Figure 4.3.12). The oxyferryl-O charges are more positive than the oxygen charge in CcP and AXP. There is no significant difference in the positive charge of the pNMH residue when compared to the pHIS.

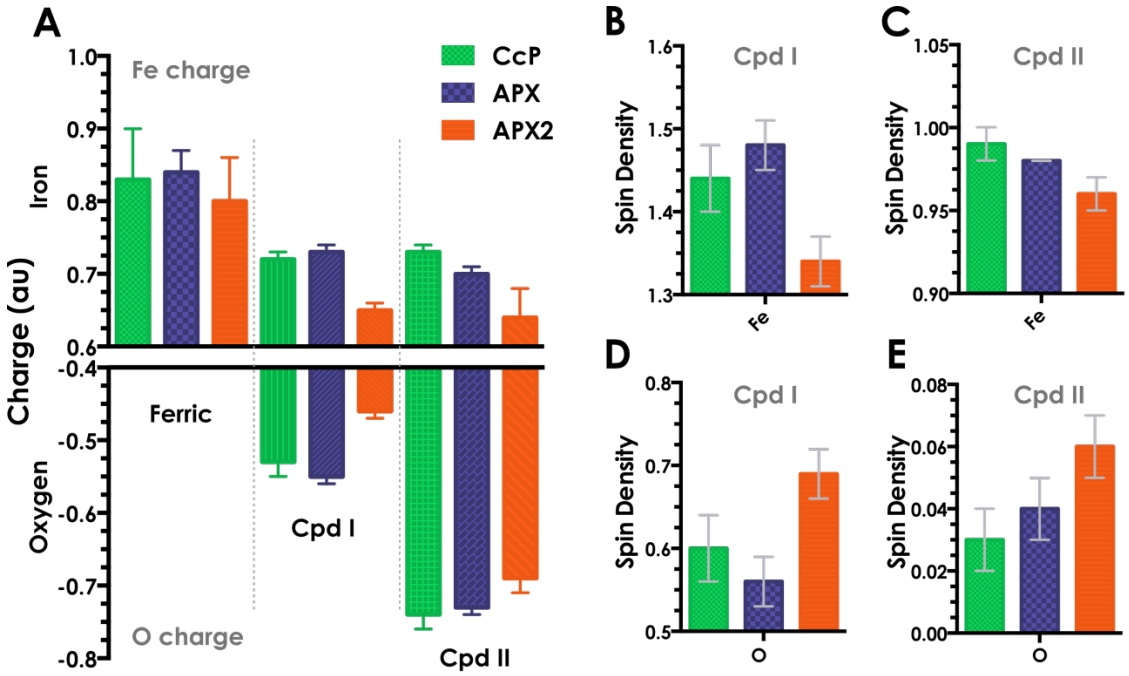


Figure 4.3.12 Atomic charges and spin density on heme Fe and oxyferryl-O. *Charges and spin densities are calculated from Mulliken population analysis and are displayed atomic units (au). Error bars show the standard deviation from the mean.*

4.4 Discussion

4.4.1 QM Cluster Models

The aim of the QM cluster model geometry optimisations was to attempt to reproduce the important interactions of the protein active site. The bond distances in the first coordination sphere around the iron are in good agreement with the experimental structures. They support the assignment of non-protonated Fe=O compound I in CcP and a protonated Fe-OH compound II in APX. The experimentally observed compound I radical was also reproduced by the cluster model, supporting the assignment of $pTRP^{\bullet+}$ radical cation in CcP [156].

In many of the cluster models, large movements of the heme during the minimisation highlighted the significance of the protein environment on constraining the porphyrin. Adding a dihedral constraint to minimise heme rotation was not sufficient to stop movement that is not possible in the natural system. In future, for cluster models of this type it may be beneficial to run calculations with a Cartesian restraint on the edge of the truncated propionates, rather than the dihedral constraint at the proximal ligation site. Expanding the calculation to a QM/MM hybrid model should also avoid these problems as these include the protein environment.

The rotation of the heme and pHIS disrupted the proximal hydrogen bonding network in almost all the optimised structures. The erroneous pHIS and heme rotations did raise the question of whether the orientation of the pHIS can control pTRP radical formation in CcP compound I. Tryptophan radicals were seen in all calculations where the pHIS N-H bond was interacting with the pASP or facing the pTRP. However, a pTRP radical was not observed in one calculation where the pHIS was rotated in the opposite direction.

In the ferric experimental structure the nearest water molecule is situated very close to the heme iron. This geometry was not favoured by either dHIP or dHID protonation cluster model. The water moves out into the distal pocket by just over 1 Å in the ferric dHID calculation and just under 1 Å in the ferric dHIP calculation. The nearest water to the oxyferryl

species in the neutron cryo-crystallography structure of compound I was assigned in an orientation that favoured interacting with the dHIS and not the oxyferryl-O. This orientation was then assumed to preclude the water-mediated mechanism of compound I reduction. The water molecule of the cluster model moves slightly to form hydrogen bonds with both the dHIS and oxyferryl-O. This challenges the mechanistic assumption, made by authors of the experimental structure, as it seems to be in an acceptable orientation for water mediated proton transfer to the oxyferryl-O.

The cluster models show a significant movement of dARG to the "in" position where it is "out" in the AXP experimental structure. This change may be the result of only one crystal water being included in the cluster model of each intermediate. The APX compound II experimental structure has two water molecules in the distal site of the heme. Further work to improve the cluster model could include additional Cartesian restraints on the heme and testing the effect of additional water molecules on the hydrogen bonding of the distal site and dARG prosition. Calculations with H_2O_2 in place of water in the distal site may show the protonation state on the ferric neutron structure (110-ferric-HID) to be able to better accommodate peroxide activation.

4.4.2 QM/MM calculations

The QM/MM calculations were used to expand the number of starting structure used in the QM cluster models as well as address some of their limitations. In the QM/MM optimisation, the MM protein environment provides steric and electrostatic context for the calculation to ameliorate unrealistic heme conformations. The calculations were performed with two QM solvent molecules and structures were set up that included substrates. Including a short MD simulation before the QM/MM geometry optimisation allowed the solvent of the distal site to fully adjust to the modified dHIS protonation and/or inclusion of substrate. The starting structure used in the calculation will bias the results to structures that are similar to it. Running MD simulations aimed to reduce this bias by broadening the conformations sampled but does not remove it.

QM/MM and QM cluster Fe=O bond length results for protonated compound I are in line with the neutron structure and resonance Raman measurements in CcP [223]. Xray structures of AXP compound I and CcP compound II are difficult to isolate. An APX compound I species can be formed through photoreduction of the ferrous-oxy intermediate (Fe^{2+}-O=O) which shows a Fe=O bond length of 1.73 Å [301]. This bond length is in line with a slightly impure Fe=O rather than a Fe-OH. The same authors used a similar method to generate a CcP compound II which reported a F-O bond length (1.82 Å) in line with a protonated hydroxyferryl (Fe^{4+}-OH) structure.

The proximity of dTRP to the heme iron and hydrogen bonds it forms with the oxyferryl-O was conserved across all compound I QM/MM calculations. It is clearly an important stabilising interaction of the oxyferryl species. The dARG was seen to prefer the '"in" conformation in both compound I and II precursor MD simulations. The MD input structures for QM/MM had the dARG closer to oxyferryl-O than after the QM/MM geometry optimisation. This may mean that the electrostatic parameters of the ferryl intermediate need to be better scaled to interact with the positive protein sidechains. In this project, this could have caused oversampling of the dARG "in" position for compound II calculations. However, experimental structures of AXP and CcP compound II show both the dARG "in" and "out" positions. Additionally, movement of the dARG "in" was also seen in the compound II cluster models. For further work the MM parameters for connecting each intermediate to the pHIS should be modified to reflect the average pHIS:NE2-Fe bond length for each intermediate. The results in this work all used the same CHARMM patch to attach the pHIS to the heme. As a result, the MD NE2-Fe bond was the same in all MD simulations.

The closest water to the oxy- and hydroxyferryl-O was seen to be highly ordered in all three enzymes. In all dHIP calculations it took up a position accepting a hydrogen bond from the dHIP and donation a hydrogen bond to the oxyferryl-O. In all dHID calculations it was in a similar position but donating a hydrogen bond to both the dHID and oxyferryl-O. As previously mentioned for the cluster models, this puts the water molecule, especially in the pHIP structures, in an acceptable position for water-mediated proton transfer to reduce compound I and II. The QM/MM calculations also produced increased hydrogen bonding of the dARG and solvent molecules around the dARG. As a result, protons coming for either the dARG or solvent coordinated by the dHIS are both plausible routes.

4.4. Discussion

The ferric dHIS QM/MM calculations took the longest to converge and had the highest failure rate. This could be because both compound I and II have a more constrained structure. The ferric dHID structures were the most dynamic in terms of solvent exchange for the distal site. This could have made choosing the two most appropriate water molecules to include in the QM region of the QM/MM calculation more difficult. Work to investigate how the number of active site waters affects geometry optimisation convergence would be beneficial for future calculations.

These calculations do not support the hypotheses that the pNMH in AXP2 improves the resilience of the enzyme by delocalising the radical on to the proximal ligand [333]. No change in spin density of the proximal ligand was observed. The Fe-O bond in APX2 compound I and II is slightly shorter than in the AXP2. There is also a notable shift in the charge and spin density between the Compound I iron and oxyferryl-O. In APX2 compound I the iron is assigned more negative charge and less spin density whilst the opposite is true for the oxyferryl-O. Cation binding in a site close to pTRP in AXP has been seen to be important in producing porphyrin, rather than tryptophan, radicals. The pNMH ligand in AXP2 achieves this without a cation bound.

MD simulations were set up where H_2O_2 was constrained to near the heme iron, weakly between 2.5 and 6 Å and then more forcibly at distances greater than 6 Å. The aim of these calculations was to understand more about how the dHIS protonation state might better accommodate peroxide activation. QM/MM optimised structures show the pHID structures were better able to bind H_2O_2 with it forming hydrogen bonds between the dHID, dARG and water molecule. In some QM/MM geometry optimisations direct deprotonation of H_2O_2 by the dHID occurred. These results seem to support a “dry” mechanism for the generation of the peroxyferric intermediate (compound 0) [5, 196].

Comparing QM/MM energies of structures that have undergone separate MD equilibration does not appear to be a particularly robust method to detect subtle changes in protonation state. Input structures had the same number of frozen and active atoms and the

QM region atoms and total atoms only differed by a single proton. However, when the final energies were balanced and compared they were more varied than anticipated and often differed by much more than is to be expected in a proton affinity calculation. The general trend of the calculations show that there is a preference for dHIP in the compound I and II species and less of a preference for dHIP in the ferric calculations. The calculations are then broadly in line with the protonation states reported in the neutron cryo-crystallography structures [5, 6].

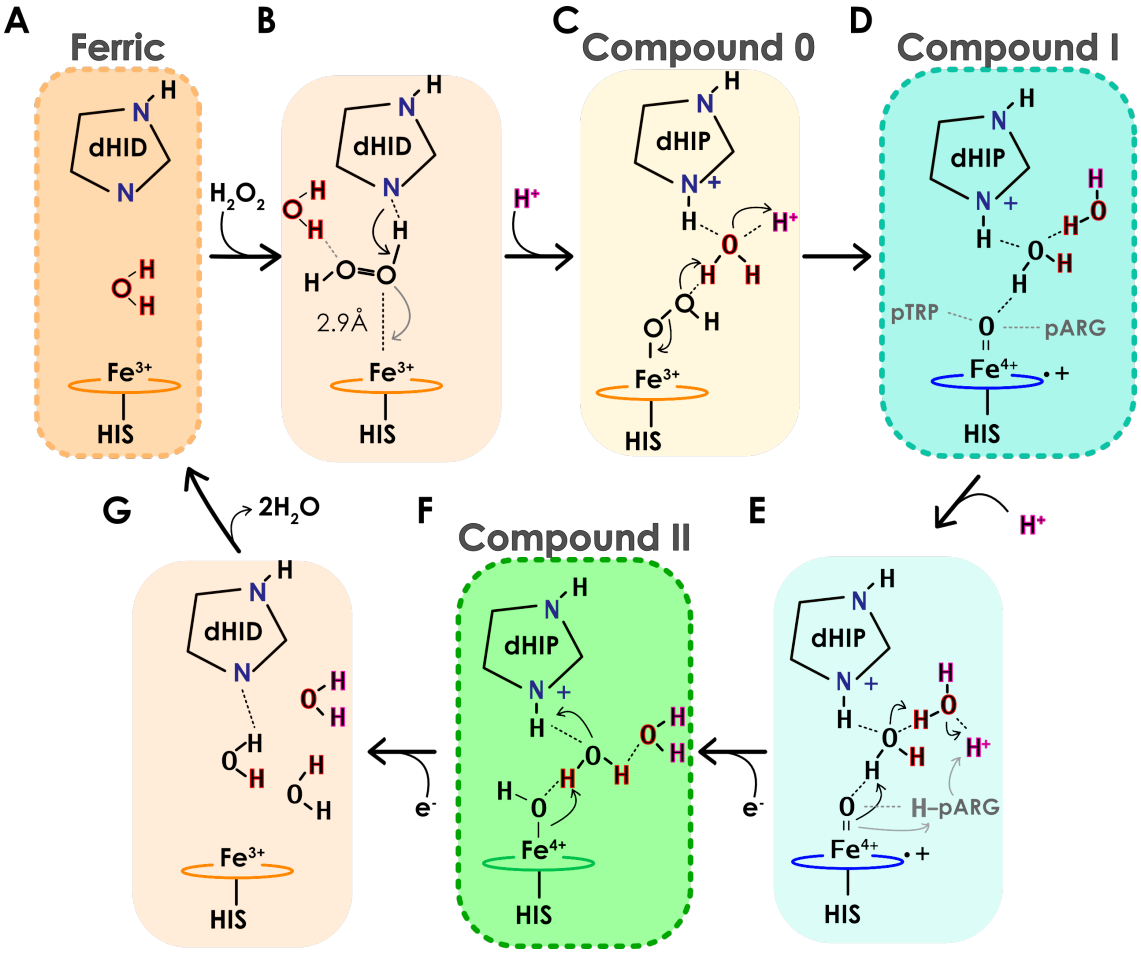


Figure 4.4.1 Suggested peroxidase mechanism. *Structures with the dashed line around them are based on the geometry optimisation of the neutron cryo-crystallography structures. The compound 0 structure was not modelled in this project. In structure* ***(B)*** *the deprotonation of H_2O_2 is proposed to occur before the formation of the peroxyferric intermediate (compound 0). In structure* ***(E)*** *and alternative protonation route via the dARG is shown in grey.*

A potentially interesting approach to further understand the role of the distal protonation could be to run the optimisation forward and backward. That is, starting with one dHIS

protonation, running a geometry optimisation, editing it to the alternative dHIS protonation, and re-running the geometry optimisation. Then a comparison could be made between the energetic and structural changes of forward and reverse optimisations. To improve the QM/MM calculations on peroxidase systems in the future it may be useful to constrain further the MD minimisations to decrease the variation it generates. This could allow specific interactions to be tested without their effect being lost in the variation generated by MD. Alternatively, increasing the number of QM/MM optimised results from different starting inputs and averaging the results could also lead to more reliable results and make it easier to unpick the effect of subtle variation on these electronically complex systems.

The results from this project support a reaction scheme such as shown in Figure 4.4.1. The wet mechanism for compound 0 formation in HRP was supported by Vidossich *et al*. with a combined MD and QM/MM analysis in 2010 [199]. They showed a lower energy barrier for water mediated deprotonation of H_2O_2 in a system were the H_2O_2 was already bonded to the heme iron. The Fe-O bond in the compound 0 intermediate used (Fe^{3+}-OHOH) was around 2.1 Å. Preliminary calculations in this project suggest that deprotonation of H_2O_2 by the pHID may occur directly at Fe-O distances near to 2.9 Å. The hydrogen bonding in the distal site of calculation of compound I and II, along with the preference for dHIS to remain protonated, would suggest a water mediated mechanism for compound I and II protonation. However, the dARG position in compound I and some of the compound II results suggest it also could play an active role in proton delivery to the ferryl intermediates.

Chapter 5

Discussion and conclusions

This project has focused on understanding the structural components necessary to produce peroxidase activity in a *de novo* protein. This was primarily achieved using the bottom-up maquette design approach. Computational modelling conducted on natural peroxidases was also performed to further our understanding of the peroxidase reactive mechanism and to lay the foundation for the same modelling techniques to be used in the *de novo* system in the future. The first two results chapters described the design and characterisation of a *de novo* peroxidase. Incorporating a *c*-type heme cofactor in the *de novo* four-helix bundle allowed the protein to be expressed as a holoenzyme *in vivo* [111]. The last design step for producing C45, the maquette peroxidase, was to remove the distal ligating histidine, leaving the final iron coordination site accessible to water. The inspiration for this change came from natural peroxidases that have a penta-coordinate heme ligation [145, 147–149]. This modification transformed an unreactive parent protein to a highly catalytically efficient thermostable *de novo* enzyme.

Tailor-made synthetic proteins that can mimic the highly specific and efficient catalysis of natural proteins are key to realising the goal of robust and sustainable catalysts for industrial biotechnology [1]. Maquettes have been designed to incorporate many natural cofactors including: heme, chlorins, flavins and quinones [54–58]. They have been used to study allosteric and charge activated conformational switching, cofactor binding, electron transfer and oxygen binding [59–62]. These results, taken with the results of this work, demonstrate the potential value the *de novo* oxidoreductases maquettes might have in bespoke synthetic pathways.

5.1 Catalysis

The covalent incorporation of *c*-type heme into the 4-helix bundle produced a protein robust enough to tolerate the removal of the distal histidine binding site. The mono-histidine maquette produced on its removal (C45), shows characteristic UV-visible spectroscopic changes associated with water bound to the distal face in the

resting state [257]. On mixing with a known electron-donating substrate, ABTS, and H_2O_2 the maquette shows impressive peroxidase activity (Figure 2.3.11). In terms of catalytic efficiency (k_{cat}/K_M), some substrates can match activity of HRP and even outperform DHP with its natural substrate (HRP activity with guaiacol and DHP activity with TBP see Table 2.3.3). C45 demonstrates that the conserved active site residues of natural peroxidase are not required to produce promiscuous peroxidase activity. Efficient peroxidase activity seems to be predominantly dependent on: creating a mono-histidine heme ligation in a protein environment that allows for the binding and activation of H_2O_2, and a heme edge that is solvent exposed to allow electron transfer from electron donating substrates.

Peroxidases generally function optimally at acidic pH's [152]. pH dependent profile of k_{cat} in C45 fits to a double pK_a curve with an optimum at pH 8.6. The difference in pH optimum is likely to be a result of not having a distal histidine residue conserved in natural peroxidases. As C45 lacks this feature at pH values under the first pK_a (pH 8), it is likely that the rate of compound 0 deprotonation is rate limiting. This inference is chiefly made by analogy to peroxygenase activity studies with a thermophilic P450 [292]. The decrease in activity above pH 8.6 points to a residue with a pK_a of 9.1 being involved in the reactive cycle. The most likely candidate is K52. In MD simulations of C45, K52 can be seen to hydrogen bond with heme propionates and is within 6.5 of the heme edge. To probe the mechanism of the C45 in greater detail mutation studies of residues that could stabilise the reactive intermediates of C45 would be interesting. K52 and Q45 should be the first residues explored as they close the heme distal site and are able to stabilise the negative charge produced during the formation of compound 0.

Similar to natural peroxidases, C45 is able to oxidase a broad range of electron-donating substrates. Many natural peroxidases do not have a single well-defined binding site for small molecule peroxidase substrates. CcP, LiP, VP, DypB and DHP have at least two small molecule binding sites, with one normally situated near the heme edge [163, 165, 166, 172, 184]. A computational binding study suggests that the same is true for C45. The most frequently sampled substrate binding site was close to the heme above L22, but the closest binding site to the heme sampled was at the heme edge. Further work to validate the binding site above L22 is needed to better understand which

binding sites are kinetically important. A first step would be to mutate L22 to perturb the binding of substrate at this site. The redox potential of the substrate, rather than a positive result in the computational binding screen, was seen to be the best indicator to predict substrate reactivity with C45. By testing a range of substrates, it was possible to infer that the redox potential of C45 compound I is close to that of 4-aminobenzoic acid (~1000 mV vs NHE).

5.2 Structure

The CD spectroscopy of the *c*-type maquette shows it is highly helical and thermostable (Figure 2.3.4). The long flexible loop regions and the lack of ordered core packing has rendered attempts at crystallisation futile. These factors combined with a highly repetitive sequence also make NMR structural assignment problematic. The inherent conformation flexibility that C45 exhibits likely aids incorporation of *c*-type heme by the Cytochrome *c* biogenesis system I (Ccm) [69]. The Rosetta folding and MD simulations of C45 suggest that it is a stable four-helical bundle but that flexibility in the second and fourth helices is likely to play an important role in binding and activating H_2O_2. Further simulation of water or H_2O_2 bound heme species could be a straight-forward method of probing what interactions may be important in allowing this flexibility.

Future work to better understand how the flexibility of the maquette effects peroxidase activity and heme incorporation will be extremely important. Appendix E shows work already underway to produce *c*-type heme maquettes with increased structural stability. *C*-type maquettes with an additional heptad repeat and a coiled-coil-C45 chimera have been designed. Extending the heptad, on the far end of the helix from the heme, could help force a homogeneous bundle conformation without any large changes to heme binding site. The coiled-coil-C45 chimera protein is made up of an idealised anti-parallel coiled-coil with the C45 heme binding portion added to each helix at one end of the bundle. Preliminary work with the extended heptad version of C45 (CXH) shows that it expresses in *E. coli*, incorporates heme and oxidises ABTS. Future work on expressing protein for NMR and crystallisation trials could produce atomistic structural information of the *c*-type maquettes.

5.3 Radical

The UV-visible intermediate observed on mixing C45 and H_2O_2 could be the maquette compound I intermediate as observed in Electron Paramagnetic Resonance (EPR) spectroscopy to produce a radical signal. Like natural peroxidases, C45 and its dW mutants can function with or without tryptophan radicals [172, 295]. One possible explanation of the results of Chapter 3 could be that, in the absence of electron-donating substrates, all three tryptophan residues on C45 are capable of producing tryptophan radicals that are observable by EPR spectroscopy (Figure 3.2.17). However, it may be the case that C45 acts in a similar way to the peroxidase activity of Cyt *c*, forming a non-specific array of protein radicals. Further work in developing the radical scavenging assay described in Section 3.3.12, needs to be explored to be able to conclusively assign the radical species [302].

Experiments with an electron-donating substrate suggest the more distant tryptophan residues are not as kinetically relevant as the closest tryptophan residue. With ABTS as an electron donor, which is predicted to primarily bind at the heme edge, having redox-active residues producing alternative radical sites reduced the efficiency of oxidation. This trend has been observed in CcP when the sequential removal of seven redox-active residues resulted in increased catalytic efficiency towards a small molecule substrate [295]. Tryptophan residues in C45 may make the enzyme more resilient (increased TON), though more work is required to develop a reliable assay for this measure. For further work, an assay with glucose oxidase and superoxide dismutase producing a steady supply of H_2O_2 without any other oxygen radicals could be the best way of testing the total turnover number in peroxidase maquettes. Reactions of glucose and glucose oxidase quickly become limited by the diffusion of oxygen into solution. With the right optimisation, mixing could then be used as a method of controlling H_2O_2 production.

5.4 QM calculations of natural peroxidases

The aim of the QM and QM/MM calculation on the natural peroxidase was to better understand how the protonation state of the distal histidine effects the distal site of natural peroxidase. These calculations were also run with a view to use that same type

of calculation on the *de novo* peroxidases in the future. Neutron cryo-crystallography structures of compound I and II, in CcP and AXP respectively, challenged the existing mechanism of proton delivery in the peroxidase mechanism [5, 6]. In these structures both the compound I and II species have the distal histidine (dHIS) doubly protonated (dHIP), which in turn means an additional proton is required in the formation on compound I. Calculations were setup for each species in pairs, one with the distal histidine singly protonated (dHID) and the other with the dHIP species. A comparison between the energies of the system and the distal hydrogen bonding networks was made to rationalise the difference between the species.

Comparing the QM cluster and QM/MM calculations energies of the dHID and dHIP calculations shows a stronger preference for dHIP in the compound I and II species. This supports the results of the neuron structures. Comparing QM/MM energies of structures that have undergone separate MD equilibration does not appear to be a particularly robust method to detect subtle changes in protonation state. When the final energies were balanced and compared they were more varied than anticipated. Further work is need to understand what part of the calculation has larger than expected energy differences. Performing QM energy calculations on the atoms of the QM region of the QM/MM calculation without geometry optimisation should be the first step in further validating these calculations. With these calculations it would then be possible to compare the structure without the energy contribution of the MM calculations, which is likely to be the source of the larger than expected energy differences.

Another approach to further validate the results of the QM/MM calculations could be to run the optimisation forwards and then backwards. That is, starting with one dHIS protonation, running a geometry optimisation, editing it to the alternative dHIS protonation, and re-running the geometry optimisation. Then a comparison could be made between the energetic and structural changes of forward and reverse optimisations.

The calculations show the ferric species is the most conformationally dynamic, making it the most difficult to model, with many of the QM/MM calculations struggling with a high number of cycles to reach the convergence criteria. In contrast, the

compound I and II species were more restrained by interactions with the oxyferryl-O. The closest distal water in both dHID and dHIP calculations formed hydrogen bonds with the oxyferryl species and the dHIS. The orientation of these interactions support a water-mediated mechanism of compound I and II deprotonation [198, 199]. Preliminary calculations in this project suggest that deprotonation of H_2O_2 by the pHID may occur directly at Fe-O distances near to 2.9 Å. Calculations on the energy of compound I formation performed by Vidossich *et al.*, showed a lower energy barrier for water mediated deprotonation of H_2O_2 in a system were the H_2O_2 was already bound to the heme (Fe-O = 2.1 Å) [199]. Further work to reproduce the energy calculation performed by Vidossich *et al.*, with peroxide starting at a geometry further from the heme (Fe-O = 2.9 Å), could produce significant insight in to compound 0 formation if this is found to be a lower energy pathway.

This project has shown the success of many years of iterative rational design to produce a *in vivo* expressed *de novo* peroxidase maquette. This flexible scaffold is extremely catalytically efficient for a *de novo* enzyme but its dynamic nature has made atomistic structural determination difficult. This project has reviewed the increasing accuracy of protein structure prediction software as well as displaying the increasing utility of existing MD and QM/MM computational methods. For flexible maquette proteins of the future, advances in computational modelling may make the need for experimental structures less pressing. These robust and highly tuneable redox proteins could be designed and folded *in silico*. Substrate binding site could be identified using MD binding simulations and the regio- and stereoselectivity of their reactions rationalised though QM/MM calculations. These functional components could then make up orthogonal synthetic pathways to sustainably produce high value compounds *in vivo*.

Bibliography

1. Channon K, Bromley EH, Woolfson DN (2008) Synthetic biology through biomolecular design and engineering. Curr Opin Struct Biol 18:491–498
2. Martínez AT, Ruiz-Dueñas FJ, Camarero S, et al (2017) Oxidoreductases on their way to industrial biotransformations. Biotechnol Adv 35:815–831
3. Lichtenstein BR, Farid T a, Kodali G, et al (2012) Engineering oxidoreductases: maquette proteins designed from scratch. Biochem Soc Trans 40:561–6
4. Armstrong CT, Watkins DW, Anderson JLR (2013) Constructing manmade enzymes for oxygen activation. Dalton Trans 42:3136–50
5. Casadei CM, Gumiero a., Metcalfe CL, et al (2014) Neutron cryo-crystallography captures the protonation state of ferryl heme in a peroxidase. Science 345:193–197
6. Kwon H, Basran J, Casadei CM, et al (2016) Direct visualisation of a Fe(IV)-OH intermediate in a heme enzyme. Nat Commun 7:1–6
7. Feynman R (1988) Richard Feynman's blackboard at time of his death. In: Calif. Inst. Technol. Image Arch. http://archives-dc.library.caltech.edu/islandora/object/ct1%3A483. Accessed 11 Sep 2017
8. Yosca TH, Behan RK, Krest CM, Onderko EL, Langston MC, Green MT (2014) Setting an upper limit on the myoglobin Iron(IV)Hydroxide pKa: Insight into axial ligand tuning in heme protein catalysis. J Am Chem Soc 136:9124–9131
9. Berman HM (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
10. Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. Sci (Washington, DC, United States) 338:1042–1046
11. Levinthal C (1969) How to Fold Graciously. In Mossbauer Spectroscopy in Biological Systems. Univ Illinois Press 22–24
12. Voet D, Voet JG, Pratt CW (2012) Fundamentals of Biochemistry: Life at the Molecular Level, 4th ed. Wiley, Hoboken, NJ
13. Pauling L, Corey RB, Branson HR (1951) The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci 37:205–211
14. Dill KA (1990) Dominant Forces in Protein Folding. Biochemistry 29:7133–7155
15. Zwanzig R, Szabo A, Bagchi B (1992) Levinthal's paradox. Proc Natl Acad Sci 89:20–22
16. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230
17. Kim PS, Baldwin RL (1982) Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding. Annu Rev Biochem 51:459–489
18. Kauzmann W (1959) Some Factors in the Interpretation of Protein Denaturation. Adv Protein Chem 14:1–63
19. Shoemaker KR, Kim PS, Brems DN, Marqusee S, York EJ, Chaiken IM, Stewart JM, Baldwin RL (1985) Nature of the charged-group effect on the stability of the C-peptide helix. Proc Natl Acad Sci U S A 82:2349–2353
20. Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding?

Trends Biochem Sci 28:18–25
21. De Prat Gay G, Ruiz-Sanz J, Neira JL, Itzhaki LS, Fersht AR (1995) Folding of a nascent polypeptide chain *in vitro*: cooperative formation of structure in a protein module. Proc Natl Acad Sci U S A 92:3683–3686
22. Pace CN, Fu H, Fryar KL, et al (2011) Contribution of hydrophobic interactions to protein stability. J Mol Biol 408:514–528
23. Uversky VN, Fink AL (2002) The chicken-egg scenario of protein folding revisited. FEBS Lett 515:79–83
24. Jackson SE, Fersht AR (1991) Folding of Chymotrypsin Inhibitor 2.1. Evidence for a Two-State Transition. Biochemistry 30:10428–10435
25. Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. Proc Natl Acad Sci U S A 91:10422–10425
26. Nölting B, Agard DA (2008) How general is the nucleation-condensation mechanism? Proteins Struct Funct Bioinforma 73:754–764
27. Rost B, Sander C (1996) Bridging the protein sequence-structure gap by structure predictions. Annu Rev Biophys Biomol Struct 25:113–136
28. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2014) Critical assessment of methods of protein structure prediction (CASP) - round x. Proteins Struct Funct Bioinforma 82:1–6
29. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248
30. Simons KT, Bonneau R, Ruczinski I, Baker D (1999) *Ab initio* protein structure prediction of CASP III targets using ROSETTA. Proteins Struct Funct Genet 37:171–176
31. Fink AL (2005) Natively unfolded proteins. Curr Opin Struct Biol 15:35–41
32. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science 302:1364–1368
33. Röthlisberger D, Khersonsky O, Wollacott AM, et al (2008) Kemp elimination catalysts by computational enzyme design. Nature 453:190–5
34. Jiang L, Althoff EA, Clemente FR, et al (2008) *De novo* computational design of retro-aldol enzymes. Science 319:1387–1391
35. Siegel JB, Zanghellini A, Lovick HM, et al (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. Science 329:309–313
36. Korendovych I V., Kulp DW, Wu Y, Cheng H, Roder H, DeGrado WF (2011) Design of a switchable eliminase. Proc Natl Acad Sci 108:6823–6827
37. Huang P-S, Boyken SE, Baker D (2016) The coming of age of *de novo* protein design. Nature 537:320–327
38. Hecht MH, Richardson JS, Richardson DC, Ogden RC (1988) *De novo* Design, Expression, and Characterization of Felix: A Four-Helix Bundle Protein of Native-Like Sequence. Science 249:884–891
39. Ho SP, DeGrado WF (1987) Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. J Am Chem Soc 109:6751–

6758
40. Regan L, DeGrado W (1988) Characterization of a helical protein designed from first principles. Science 241:976–978
41. Padmanabhan S, Marqusee S, Ridgeway T, Laue TM, Baldwin RL (1990) Relative helix-forming tendencies of nonpolar amino acids. Nature 344:268–270
42. Kim CA, Berg JM (1993) Thermodynamic β -sheet propensities measured using a zinc-finger host peptide. Nature 362:267–270
43. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685
44. Woolfson DN (2005) The Design of Coiled-Coil Structures and Assemblies. In: Adv. Protein Chem. pp 79–112
45. Crick FHC (1953) The packing of α-helices: simple coiled-coils. Acta Crystallogr 6:689–697
46. Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. Nature 371:80–83
47. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346:485–488
48. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL (1994) Design and synthesis of multi-haem proteins. Nature 368:425–32
49. Chu R, Pei W, Takei J, Bai Y (2002) Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. Biochemistry 41:7998–8003
50. Laidig KE, Daggett V (1996) Molecular dynamics simulations of apocytochrome *b*562- The highly ordered limit of molten globules. Fold Des 1:335–346
51. Presnell SR, Cohen FE (1989) Topological distribution of four-alpha-helix bundles. Proc Natl Acad Sci U S A 86:6592–6596
52. Chou PY, Fasman GD (1978) Empirical Predictions of Protein Conformation. Annu Rev Biochem 47:251–276
53. Lawrence MS, Phillips KJ, Liu DR (2007) Supercharging proteins can impart unusual resilience. J Am Chem Soc 129:10110–10112
54. Gibney BR, Isogai Y, Rabanal F, Reddy KS, Grosset AM, Moser CC, Dutton PL (2000) Self-assembly of heme A and heme B in a designed four-helix bundle: Implications for a Cytochrome *c* oxidase maquette. Biochemistry 39:11041–11049
55. Razeghifard MR, Wydrzynski T (2003) Binding of Zn-chlorin to a synthetic four-helix bundle peptide through histidine ligation. Biochemistry 42:1024–1030
56. Sharp RE, Moser CC, Rabanal F, Dutton PL (1998) Design, synthesis, and characterization of a photoactivatable flavocytochrome molecular maquette. Proc Natl Acad Sci U S A 95:10465–70
57. Rabanal F, Gibney BR, DeGrado WF, Moser CC, Leslie Dutton P (1996) Engineering photosynthesis: synthetic redox proteins. Inorganica Chim Acta 243:213–218
58. Farid T a, Kodali G, Solomon L a, et al (2013) Elementary tetrahelical protein design for diverse oxidoreductase functions. Nat Chem Biol 9:826–33

59. Grosset AM, Gibney BR, Rabanal F, Moser CC, Dutton PL (2001) Proof of Principle in a *de novo* Designed Protein Maquette: An Allosterically Regulated, Charge-Activated Conformational Switch in a Tetra-α-Helix Bundle. Biochemistry 40:5474–5487
60. Solomon LA, Kodali G, Moser CC, Dutton PL (2014) Engineering the Assembly of Heme Cofactors in Man-Made Proteins. J Am Chem Soc 136:3192–3199
61. Shifman JM, Moser CC, Kalsbeck WA, Bocian DF, Dutton PL (1998) Functionalized *de novo* Designed Proteins: Mechanism of Proton Coupling to Oxidation/Reduction in Heme Protein Maquettes. Biochemistry 37:16815–16827
62. Koder RL, Anderson JLR, Solomon L a, Reddy KS, Moser CC, Dutton PL (2009) Design and engineering of an O(2) transport protein. Nature 458:305–9
63. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL (1994) Design and synthesis of multi-haem proteins. Nature 368:425–32
64. Gibney BR, Rabanal F, Skalicky JJ, Wand AJ, Dutton PL, Uni V, Pennsyl V (1999) Iterative Protein Redesign. 4952–4960
65. Huang SS, Gibney BR, Stayrook SE, Leslie Dutton P, Lewis M (2003) X-ray Structure of a Maquette Scaffold. J Mol Biol 326:1219–1225
66. Huang SS, Koder RL, Lewis M, Wand a J, Dutton PL (2004) The HP-1 maquette: from an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. Proc Natl Acad Sci U S A 101:5536–41
67. Shikama K (1998) The Molecular Mechanism of Autoxidation for Myoglobin and Hemoglobin: A Venerable Puzzle. Chem Rev 98:1357–1374
68. Negron C, Fufezan C, Koder RL (2009) Geometric constraints for porphyrin binding in helical protein binding sites. Proteins 74:400–16
69. Watkins DW, Armstrong CT, Beesley JL, Marsh JE, Jenkins JMX, Sessions RB, Mann S, Ross Anderson JL (2016) A suite of *de novo c*-type cytochromes for functional oxidoreductase engineering. Biochim Biophys Acta - Bioenerg 1857:493–502
70. Watkins DW, Jenkins JMX, Grayson KJ, et al (2017) Construction and *in vivo* assembly of a catalytically proficient and hyperthermostable *de novo* enzyme. Nat Commun 8:358
71. Watkins DW, Armstrong CT, Anderson JLR (2014) *De novo* protein components for oxidoreductase assembly and biological integration. Curr Opin Chem Biol 19:90–8
72. Faraone-Mennella J, Tezcan FA, Gray HB, Winkler JR (2006) Stability and folding kinetics of structurally characterized Cytochrome *c*-b562. Biochemistry 45:10504–10511
73. Dumont ME, Campbell GA, Coring AF (1994) Noncovalent Binding of Heme Induces a Compact Apocytochrome *c* Structure. Biochemistry 33:7368–7378
74. Discher BM, Noy D, Strzalka J, Ye S, Moser CC, Lear JD, Blasie JK, Dutton PL (2005) Design of amphiphilic protein maquettes: Controlling assembly, membrane insertion, and cofactor interactions. Biochemistry 44:12329–12343
75. Shifman JM, Gibney BR, Sharp RE, Dutton PL (2000) Heme redox potential control in *de novo* designed four-α-helix bundle proteins. Biochemistry 39:14813–

14821
76. Schwarz G, Mendel RR, Ribbe MW (2009) Molybdenum cofactors, enzymes and pathways. Nature 460:839–847
77. Hörnberg A, Logan DT, Marklund SL, Oliveberg M (2007) The Coupling between Disulphide Status, Metallation and Dimer Interface Strength in Cu/Zn Superoxide Dismutase. J Mol Biol 365:333–342
78. Fukuyama K, Okada T, Kakuta Y, Takahashi Y (2002) Atomic resolution structures of oxidized [4Fe-4S] ferredoxin from Bacillus thermoproteolyticus in two crystal forms: Systematic distortion of [4Fe-4S] cluster in the protein. J Mol Biol 315:1155–1166
79. Knapp S, Kardinahl S, Hellgren N, Tibbelin G, Schäfer G, Ladenstein R (1999) Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon Sulfolobus acidocaldarius at 2.2 Å resolution. J Mol Biol 285:689–702
80. Marcus RA (1956) On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. J Phys Chem 24:966–978
81. Hopfield JJ (1974) Electron Transfer Between Biological Molecules by Thermally Activated Tunneling. Proc Natl Acad Sci 71:3640–3644
82. Moser CC, Farid TA, Chobot SE, Dutton PL (2006) Electron tunneling chains of mitochondria. Biochim Biophys Acta - Bioenerg 1757:1096–1109
83. Moser CC, Anderson JLR, Dutton PL (2010) Guidelines for tunneling in enzymes. Biochim Biophys Acta 1797:1573–86
84. Page CC, Moser CC, Chen X, Dutton PL (1999) Natural engineering principles of electron tunnelling in biological oxidation-reduction. Nature 402:47–52
85. Macedo AL, Besson S, Moreno C, Fauque G, Moura JJG, Moura I (1996) Characterization of a 7Fe ferredoxin isolated from the marine denitrifier pseudomonas nautica strain 617: Spectroscopic and electrochemical studies. Biochem Biophys Res Commun 229:524–530
86. Ishikita H, Knapp EW (2003) Redox Potential of Quinones in Both Electron Transfer Branches of Photosystem I. J Biol Chem 278:52002–52011
87. Deistung J, Thorneley RN (1986) Electron transfer to nitrogenase. Characterization of flavodoxin from Azotobacter chroococcum and comparison of its redox potentials with those of flavodoxins from Azotobacter vinelandii and Klebsiella pneumoniae (nifF-gene product). Biochem J 239:69–75
88. Robertson DE, Prince RC, Bowyer JR, Matsuura K, Dutton PL, Ohnisho T (1984) Thermodynamic Properties of the semiquinone and its binding site in the ubiquinol-cytochrome *c* (c_2) oxireductase of respiratory and photosynthetic systems. J od Biol Chem 259:1758–1763
89. Ashur I, Brandis A, Greenwald M, Vakrat-Haglili Y, Rosenbach-Belkin V, Scheer H, Scherz A (2003) Control of redox transitions and oxygen species binding in Mn centers by biologically significant ligands; model studies with [Mn]-bacteriochlorophyll a. J Am Chem Soc 125:8852–8861
90. Machonkin TE, Zhang HH, Hedman B, Hodgson KO, Solomon EI (1998) Spectroscopic and magnetic studies of human ceruloplasmin: Identification of a redox-inactive reduced type 1 copper site. Biochemistry 37:9570–9578
91. Crofts AR (2004) Proton-coupled electron transfer at the Qo-site of the bc1 complex controls the rate of ubihydroquinone oxidation. Biochim Biophys Acta

1655:77–92
92. Heinnickel M, Golbeck JH (2007) Heliobacterial photosynthesis. Photosynth Res 92:35–53
93. Ghanem M, Gadda G (2006) Effects of reversing the protein positive charge in the proximity of the flavin N(1) locus of choline oxidase. Biochemistry 45:3437–3447
94. Nouailler M, Bruscella P, Lojou E, Lebrun R, Bonnefoy V, Guerlesquin F (2006) Structural analysis of the HiPIP from the acidophilic bacteria: Acidithiobacillus ferrooxidans. Extremophiles 10:191–198
95. Ishikita H, Knapp EW (2006) Function of redox-active tyrosine in photosystem II. Biophys J 90:3886–3896
96. Ishikita H, Loll B, Biesiadka J, Saenger W, Knapp EW (2005) Redox Potentials of Chlorophylls in the Photosystem II Reaction Center. Biochemistry 44:4118–4124
97. Izadi N, Henry Y, Haladjian J, Goldberg ME, Wandersman C, Delepierre M, Lecroisey A (1997) Purification and characterization of an extracellular heme-binding protein, HasA, involved in heme iron acquisition. Biochemistry 36:7050–7057
98. Christenson A, Dimcheva N, Ferapontova EE, et al (2004) Direct electron transfer between ligninolytic redox enzymes and electrodes. Electroanalysis 16:1074–1092
99. Nishida CR, Ortiz De Montellano PR (2005) Thermophilic cytochrome P450 enzymes. Biochem Biophys Res Commun 338:437–445
100. Mondal MS, Fuller HA, Armstrong FA (1996) Direct Measurement of the Reduction Potential of Catalytically Active Cytochrome *c* Peroxidase Compound I: Voltammetric Detection of a Reversible, Cooperative Two-Electron Transfer Reaction. J Am Chem Soc 118:263–264
101. Efimov I, Papadopoulou ND, McLean KJ, Badyal SK, Macdonald IK, Munro AW, Moody PCE, Raven EL (2007) The redox properties of ascorbate peroxidase. Biochemistry 46:8017–8023
102. Dewilde S, Kiger L, Burmester T, Hankeln T, Baudin-Creuza V, Aerts T, Marden MC, Caubergs R, Moens L (2001) Biochemical Characterization and Ligand Binding Properties of Neuroglobin, a Novel Member of the Globin Family. J Biol Chem 276:38949–38955
103. Oyadomari M, Shinohara H, Johjima T, Wariishi H, Tanaka H (2003) Electrochemical characterization of lignin peroxidase from the white-rot basidiomycete Phanerochaete chrysosporium. J Mol Catal B Enzym 21:291–297
104. Hayashi T, Murata D, Makino M, Sugimoto H, Matsuo T, Sato H, Shiro Y, Hisaeda Y (2006) Crystal structure and peroxidase activity of myoglobin reconstituted with iron porphycene. Inorg Chem 45:10530–10536
105. D'Antonio J, D'Antonio EL, Thompson MK, Bowden EF, Franzen S, Smirnova T, Ghiladi RA (2010) Spectroscopic and mechanistic investigations of dehaloperoxidase B from amphitrite ornata. Biochemistry 49:6600–6616
106. Forquer I, Covian R, Bowman MK, Trumpower BL, Kramer DM (2006) Similar transition states mediate the Q-cycle and superoxide production by the cytochrome bc1complex. J Biol Chem 281:38459–38465
107. Arciero DM, Hooper AB (1994) A di-heme Cytochrome *c* peroxidase from

Nitrosomonas europaea catalytically active in both the oxidized and half-reduced states. J Biol Chem 269:11878–11886
108. Purcell WL, Erman JE (1976) Cytochrome *c* peroxidase catalyzed oxidations of substitution inert iron(II) complexes. J Am Chem Soc 98:7033–7037
109. De Montellano PRO (2005) Cytochrome P450: structure, mechanism, and biochemistry. Springer Science & Business Media
110. Zhang L, Andersen EME, Khajo A, Magliozzo RS, Koder RL (2013) Dynamic factors affecting gaseous ligand binding in an artificial oxygen transport protein. Biochemistry 52:447–455
111. Anderson JLRR, Armstrong CT, Kodali G, et al (2014) Constructing a man-made *c*-type cytochrome maquette *in vivo*: electron transfer, oxygen transport and conversion to a photoactive light harvesting maquette. Chem Sci 5:507–514
112. WÖHRLE D, SCHNURPFEIL G (2003) Porphyrins and Phthalocyanines in Macromolecules. In: Porphyr. Handb. Academic Press, San Diego, Calif, pp 177–246
113. Green MT (2009) C{single bond}H bond activation in heme proteins: the role of thiolate ligation in cytochrome P450. Curr Opin Chem Biol 13:84–88
114. Kirkman HN, Gaetani GF (2007) Mammalian catalase: a venerable enzyme with new mysteries. Trends Biochem Sci 32:44–50
115. Mirkin N, Jaconcic J, Stojanoff V, Moreno A (2007) High resolution X-ray crystallographic structure of bovine heart Cytochrome *c* and its application to the design of an electron transfer biosensor. Proteins Struct Funct Bioinforma 70:83–92
116. Guallar V, Olsen B (2006) The role of the heme propionates in heme biochemistry. J Inorg Biochem 100:755–760
117. Hargrove MS, Wilkinson AJ, Olson JS (1996) Structural factors governing hemin dissociation from metmyoglobin. Biochemistry 35:11300–11309
118. Weinert EE, Phillips-Piro CM, Marletta MA (2013) Porphyrin π-stacking in a heme protein scaffold tunes gas ligand affinity. J Inorg Biochem 127:7–12
119. Poulos TL (2014) Heme enzyme structure and function. Chem Rev 114:3919–3962
120. Bach A Peroxydasen als specifisch wirkende Enzyme. Hrn. R. Chodat zur Antwort. Berichte der Dtsch Chem Gesellschaft 39:3329–3331
121. Altschul AM, Abrams R, Hogness TR (1940) Cytochrome *c* Peroxidase. J Biol Chem 3:777–794
122. Klingenberg M (1958) Pigments of rat liver microsomes. Arch Biochem Biophys 75:376–386
123. Garfinkel D (1958) Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. Arch Biochem Biophys 77:493–509
124. Omura, Tsuneo; Sato R (1964) The Carbon Monoxide-biding pigment of Liver Microsomes. J. Biol. Chem.
125. Takahashi S, Yeo YS, Zhao Y, O'Maille PE, Greenhagen BT, Noel JP, Coates RM, Chappell J (2007) Functional Characterization of Premnaspirodiene Oxygenase, a Cytochrome P450 Catalyzing Regio- And Stereo-specific Hydroxylations of Diverse Sesquiterpene Substrates. J Biol Chem 282:31744–31754

126. Bach RD, Mintcheva I, Estévez CM, Schlegel HB (1995) Theoretical Model for an Alternate Mechanism for the Cytochrome P-450 Hydroxylation of Quadricyclane. J Am Chem Soc 117:10121–10122
127. Estabrook RW, Cooper DY, Rosenthal O (1963) The light reversible carbon monoxide inhibition of the steroid C21-hydroxylase system of the adrenal cortex. Biochem z 338:741–755
128. Watanabe M, Kikuchi H, Sagami I, Fujii H, Ohmachi T (1989) Role of cytochrome P-450 in chemical toxicity and carcinogenesis. J Toxicol Pathol 2:241–251
129. Obach RS (2005) The Utility of *in vitro* Cytochrome P450 Inhibition Data in the Prediction of Drug-Drug Interactions. J Pharmacol Exp Ther 316:336–348
130. Chefson A, Auclair K (2006) Progress towards the easier use of P450 enzymes. Mol Biosyst 2:462–469
131. Brandenberg OF, Fasan R, Arnold FH (2017) Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. Curr Opin Biotechnol 47:102–111
132. Wang ZJ, Renata H, Peck NE, Farwell CC, Coelho PS, Arnold FH (2014) Improved cyclopropanation activity of histidine-ligated cytochromeP450 enables the enantioselective formal synthesis of levomilnacipran. Angew Chemie - Int Ed 53:6810–6813
133. Arnold F (2018) Arnold Group Patents. In: Calif. Inst. Technol. http://fhalab.caltech.edu/?page_id=128. Accessed 2 Sep 2018
134. Lewis DFV, Hlavica P (2000) Interactions between redox partners in various cytochrome P450 systems: Functional and structural aspects. Biochim Biophys Acta - Bioenerg 1460:353–374
135. Isin EM, Guengerich FP (2008) Substrate binding to cytochromes P450. Anal Bioanal Chem 392:1019–1030
136. Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. J Mol Biol 195:687–700
137. Lee YT, Wilson RF, Rupniewski I, Goodin DB (2010) P450cam visits an open conformation in the absence of substrate. Biochemistry 49:3412–3419
138. Hrycay EG, Gustafsson J-åke, Ingelman-Sundberg M, Ernster L (1975) Sodium periodate, sodium chlorite, organic hydroperoxides, and H2O2 as hydroxylating agents in steroid hydroxylation reactions catalyzed by partially purified cytochrome P-450. Biochem Biophys Res Commun 66:209–216
139. Groves JT, McClusky GA (1976) Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. J Am Chem Soc 98:859–861
140. Wang Y, Lan D, Durrani R, Hollmann F (2017) Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? Curr Opin Chem Biol 37:1–9
141. Holtmann D, Hollmann F (2016) The Oxygen Dilemma: A Severe Challenge for the Application of Monooxygenases? ChemBioChem 17:1391–1398
142. Jung ST, Lauchli R, Arnold FH (2011) Cytochrome P450: taming a wild type enzyme. Curr Opin Biotechnol 22:809–817
143. Hofrichter M, Ullrich R (2014) Oxidations catalyzed by fungal peroxygenases. Curr Opin Chem Biol 19:116–125
144. Sono M, Roach MP, Coulter ED, Dawson JH (1996) Heme-Containing

Oxygenases. Chem Rev 96:2841–2888
145. Poulos TL, Freer ST, Alden RA, Edwards SL, Skogland U, Takio K, Eriksson B, Xuong N, Yonetani T, Kraut J (1980) The crystal structure of Cytochrome *c* peroxidase. J Biol Chem 255:575–580
146. Zeng J, Fenna RE (1992) X-ray crystal structure of canine myeloperoxidase at 3 Å resolution. J Mol Biol 226:185–207
147. Poulos TL, Edwards SL, Wariishi H, Gold MH (1993) Crystallographic refinement of lignin peroxidase at 2 Å. J Biol Chem 268:4429–4440
148. Patterson WR, Poulos TL (1995) Crystal structure of recombinant pea cytosolic ascorbate peroxidase. Biochemistry 34:4331–4341
149. Gajhede M, Schuller DJ, Henriksen A, Smith AT, Poulos TL (1997) Crystal structure of horseradish peroxidase C at 2.15 Å resolution. Nat Struct Biol 4:1032–1038
150. La Mar GN, de Ropp JS, Smith KM, Langry KC (1980) Proton nuclear magnetic resonance study of the electronic and molecular structure of the heme crevice in horseradish peroxidase. J Biol Chem 255:6646–6652
151. Kwon H, Moody PCE, Raven EL (2016) Chapter 3 Understanding the Reactivity and Interactions of Peroxidases with Substrates. In: Raven E, Dunford B (eds) Heme Peroxidases. The Royal Society of Chemistry, Cambridge, UK, pp 47–60
152. Raven E, Dunford B (eds) (2015) Heme Peroxidases. The Royal Society of Chemistry, Cambridge, UK. doi: 10.1039/9781782622628
153. Pelletier H, Kraut J (1992) Crystal structure of a complex between electron transfer partners, Cytochrome *c* peroxidase and Cytochrome *c*. Science 258:1748–1755
154. Volkov AN, Nicholls P, Worrall J a R (2011) The complex of Cytochrome *c* and Cytochrome *c* peroxidase: the end of the road? Biochim Biophys Acta 1807:1482–503
155. Yonetani T, Schleyer H (1966) Studies on Cytochrome *c* peroxidase. VII. Electron paramagnetic resonance absorptions of the enzyme and complex ES in dissolved and crystalline forms. J Biol Chem 241:3240–3243
156. Sivaraja M, Goodin D, Smith M, Hoffman B (1989) Identification by ENDOR of Trp191 as the free-radical site in Cytochrome *c* peroxidase compound ES. Science 245:738–740
157. Leigh JS, Maltempo MM, Ohlsson PI, Paul KG (1975) Optical, NMR and EPR properties of horseradish peroxidase and its donor complexes. FEBS Lett 51:304–308
158. Burns PS, Williams RJP, Wright PE (1975) Conformational studies of peroxidase-substrate complexes. Structure of the indolepropionic acid-horseradish peroxidase complex. J Chem Soc Chem Commun 795–796
159. Ator MA, David SK, Ortiz de Montellano PR (1987) Structure and catalytic mechanism of horseradish peroxidase. Regiospecific meso alkylation of the prosthetic heme group by alkylhydrazines. J Biol Chem 262:14954–14960
160. Ortiz de Montellano PR, Choe YS, DePillis G, Catalano CE (1987) Structure-mechanism relationships in hemoproteins. Oxygenations catalyzed by chloroperoxidase and horseradish peroxidase. J Biol Chem 262:11641–11646
161. Sundaramoorthy M, Kishi K, Gold MH, Poulos TL (1997) Crystal structures of

substrate binding site mutants of manganese peroxidase. J Biol Chem 272:17574–17580
162. Sharp KH, Mewies M, Moody PCE, Raven EL (2003) Crystal structure of the ascorbate peroxidase–ascorbate complex. Nat Struct Biol 10:303–307
163. Murphy EJ, Metcalfe CL, Nnamchi C, Moody PCE, Raven EL (2012) Crystal structure of guaiacol and phenol bound to a heme peroxidase. FEBS J 279:1632–1639
164. Miner KD, Pfister TD, Hosseinzadeh P, Karaduman N, Donald LJ, Loewen PC, Lu Y, Ivancich A (2014) Identifying the elusive sites of tyrosyl radicals in Cytochrome *c* peroxidase: Implications for oxidation of substrates bound at a site remote from the heme. Biochemistry 53:3781–3789
165. Ruiz-Dueñas FJ, Morales M, García E, Miki Y, Martínez MJ, Martínez AT (2009) Substrate oxidation sites in versatile peroxidase and other basidiomycete peroxidases. J Exp Bot 60:441–452
166. Strittmatter E, Liers C, Ullrich R, Wachter S, Hofrichter M, Plattner DA, Piontek K (2013) First crystal structure of a fungal high-redox potential dye-decolorizing peroxidase substrate interaction sites and long-range electron transfer. J Biol Chem 288:4095–4102
167. Hammel KE, Cullen D (2008) Role of fungal peroxidases in biological ligninolysis. Curr Opin Plant Biol 11:349–355
168. Lundquist K, Kirk TK (1978) *De novo* synthesis and decomposition of veratryl alcohol by a lignin-degrading basidiomycete. Phytochemistry 17:1676
169. Kersten PJ, Kalyanaraman B, Hammel KE, Reinhammar B, Kirk TK (1990) Comparison of lignin peroxidase, horseradish peroxidase and laccase in the oxidation of methoxybenzenes. Biochem J 268:475–480
170. Cavalieri EL, Rogan EG (1990) Radical cations in aromatic hydrocarbon carcinogenesis. Free Radic Res 11:77–87
171. Hammel KE, Kalyanaraman B, Kirk TK (1986) Oxidation of polycyclic aromatic hydrocarbons and dibenzo[p]dioxins by Phanerochaete chrysosporium ligninase. J Biol Chem 261:16948–16952
172. Choinowski T, Blodig W, Winterhalter KH, Piontek K (1999) The crystal structure of lignin peroxidase at 1.70 Å resolution reveals a hydroxy group on the C(β) of tryptophan 171: A novel radical site formed during the redox cycle. J Mol Biol 286:809–827
173. Welinder KG (1992) Superfamily of plant, fungal and bacterial peroxidases. Curr Opin Struct Biol 2:388–393
174. Sugano Y (2009) DyP-type peroxidases comprise a novel heme peroxidase family. Cell Mol Life Sci 66:1387–1403
175. Sugano Y, Muramatsu R, Ichiyanagi A, Sato T, Shoda M (2007) DyP, a Unique Dye-decolorizing Peroxidase, Represents a Novel Heme Peroxidase Family. J Biol Chem 282:36652–36658
176. Liers C, Bobeth C, Pecyna M, Ullrich R, Hofrichter M (2010) DyP-like peroxidases of the jelly fungus Auricularia auricula-judae oxidize nonphenolic lignin model compounds and high-redox potential dyes. Appl Microbiol Biotechnol 85:1869–1879
177. Brown ME, Barros T, Chang MCY (2012) Identification and Characterization of

a Multifunctional Dye Peroxidase from a Lignin-Reactive Bacterium. ACS Chem Biol 7:2074–2081
178. Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TDH (2011) Identification of DypB from *rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry 50:5096–5107
179. Hardison R (1998) Hemoglobins From Bacteria To Man : Evolution of Different Patterns of Gene Expression. J Exp Biol 1117:1099–1117
180. Nienhaus K, Deng P, Belyea J, Franzen S, Nienhaus GU (2006) Spectroscopic study of substrate binding to the carbonmonoxy form of dehaloperoxidase from amphitrite ornata. J Phys Chem B 110:13264–13276
181. Le P, Zhao J, Franzen S (2014) Correlation of heme binding affinity and enzyme kinetics of dehaloperoxidase. Biochemistry 53:6863–6877
182. Nastri F, Lista L, Ringhieri P, Vitale R, Faiella M, Andreozzi C, Travascio P, Maglio O, Lombardi A, Pavone V (2011) A heme-peptide metalloenzyme mimetic with natural peroxidase-like activity. Chem - A Eur J 17:4444–4453
183. Franzen S, Thompson MK, Ghiladi RA (2012) The dehaloperoxidase paradox. Biochim Biophys Acta - Proteins Proteomics 1824:578–588
184. Wang C, Lovelace LL, Sun S, Dawson JH, Lebioda L (2013) Complexes of dual-function hemoglobin/dehaloperoxidase with substrate 2,4,6-trichlorophenol are inhibitory and indicate binding of halophenol to compound I. Biochemistry 52:6203–6210
185. Plummer A, Thompson MK, Franzen S (2013) Role of Polarity of the Distal Pocket in the Control of Inhibitor Binding in Dehaloperoxidase-Hemoglobin. Biochemistry 52:2218–2227
186. Zhao J, De Serrano V, Zhao J, Le P, Franzen S (2013) Structural and kinetic study of an internal substrate binding site in dehaloperoxidase-hemoglobin a from amphitrite ornata. Biochemistry 52:2427–2439
187. D'Antonio EL, Bowden EF, Franzen S (2012) Thin-layer spectroelectrochemistry of the Fe(III)/Fe(II) redox reaction of dehaloperoxidase-hemoglobin. J Electroanal Chem 668:37–43
188. Koch C, Sures B (2018) Environmental concentrations and toxicology of 2,4,6-tribromophenol (TBP). Environ Pollut 233:706–713
189. Leonetti C, Butt CM, Hoffman K, Hammel SC, Miranda ML, Stapleton HM (2016) Brominated flame retardants in placental tissues: associations with infant sex and thyroid hormone endpoints. Environ Heal 15:113
190. Chance B (1943) The kinetics of the enzyme-substrate compound of peroxidase. J Biol Chem 151:553–577
191. Dunford HB, Hasinoff BB (1970) Kinetics of the oxidation of ferrocyanide by horseradish peroxidase compounds I and II. Biochemistry 9:4930–4939
192. Yamada H, Yamazaki I (1974) Proton balance in conversions between five oxidation-reduction states of horseradish peroxidase. Arch Biochem Biophys 165:728–738
193. Rodriguez-Lopez JN, Smith AT, Thorneley RNF (1996) Role of arginine 38 in horseradish peroxidase: A critical residue for substrate binding and catalysis. J Biol Chem 271:4023–4030
194. Poulos TL, Kraut J (1980) The stereochemistry of peroxidase catalysis. J Biol

Chem 255:8199–8205
195. Erman JE, Vitello LB, Miller MA, Shaw A, Brown KA, Kraut J (1993) Histidine 52 is a critical residue for rapid formation of Cytochrome *c* peroxidase compound I. Biochemistry 32:9798–9806
196. Derat E, Shaik S (2006) The Poulos−Kraut Mechanism of Compound I Formation in Horseradish Peroxidase: A QM/MM Study. J Phys Chem B 110:10526–10533
197. Baek HK, Van Wart HE (1992) Elementary Steps in the Reaction of Horseradish Peroxidase with Several Peroxides: Kinetics and Thermodynamics of Formation of Compound 0 and Compound I. J Am Chem Soc 114:718–725
198. Derat E, Shaik S, Rovira C, Vidossich P, Alfonso-Prieto M (2007) The effect of a water molecule on the mechanism of formation of compound 0 in horseradish peroxidase. J Am Chem Soc 129:6346–6347
199. Vidossich P, Fiorin G, Alfonso-Prieto M, Derat E, Shaik S, Rovira C (2010) On the Role of Water in Peroxidase Catalysis: A Theoretical Investigation of HRP Compound I Formation. J Phys Chem B 114:5161–5169
200. Jones P, Dunford HB (2005) The mechanism of compound I formation revisited. J Inorg Biochem 99:2292–2298
201. Chreifi G, Baxter EL, Doukov T, Cohen AE, McPhillips SE, Song J, Meharenna YT, Soltis SM, Poulos TL (2016) Crystal structure of the pristine peroxidase ferryl center and its relevance to proton-coupled electron transfer. Proc Natl Acad Sci 113:1226–1231
202. Terner J, Palaniappan V, Gold A, Weiss R, Fitzgerald MM, Sullivan AM, Hosten CM (2006) Resonance Raman spectroscopy of oxoiron(IV) porphyrin π-cation radical and oxoiron(IV) hemes in peroxidase intermediates. J Inorg Biochem 100:480–501
203. Vitello LB, Erman JE, Miller MA, Wang J, Kraut J (1993) Effect of Arginine-48 Replacement on the Reaction between Cytochrome *c* Peroxidase and Hydrogen Peroxide. Biochemistry 32:9807–9818
204. Bonagura CA, Bhaskar B, Shimizu H, Li H, Sundaramoorthy M, McRee DE, Goodin DB, Poulos TL (2003) High-resolution crystal structures and spectroscopy of native and Compound I Cytochrome *c* peroxidase. Biochemistry 42:5600–5608
205. Groves JT, Boaz NC (2014) Fishing for peroxidase protons. Science 345:142–143
206. Dolphin D, Forman A, Borg DC, Fajer J, Felton RH (1971) Compounds I of Catalase and Horse Radish Peroxidase: pi-Cation Radicals. Proc Natl Acad Sci 68:614–618
207. Mauro JM, Fishel LA, Hazzard JT, Meyer TE, Tollin G, Cusanovich MA, Kraut J (1988) Tryptophan-191→ Phenylalanine, a Proximal-side Mutation in Yeast Cytochrome *c* Peroxidase that Strongly Affects the Kinetics of Ferrocytochrome *c* Oxidation. Biochemistry 27:6243–6256
208. Patterson WR, Poulos TL, Goodin DB (1995) Identification of a Porphyrin π Cation Radical in Ascorbate Peroxidase Compound I. Biochemistry 34:4342–4345
209. Marquez LA, Quitoriano M, Zilinskas BA, Dunford HB (1996) Kinetic and spectral properties of pea cytosolic ascorbate peroxidase. FEBS Lett 389:153–156

210. Fitzgerald MM, Churchill MJ, McRee DE, Goodin DB (1994) Small Molecule Binding to an Artificially Created Cavity at the Active Site of Cytochrome *c* Peroxidase. Biochemistry 33:3807–3818
211. Jensen GM, Bunte SW, Warshel a, Goodin DB (1998) Energetics of Cation Radical Formation at the Proximal Active Site Tryptophan of Cytochrome *c* Peroxidase and Ascorbate Peroxidase. J Phys Chem B 102:8221–8228
212. Schuller DJ, Ban N, van Huystee RB, McPherson A, Poulos TL (1996) The crystal structure of peanut peroxidase. Structure 4:311–321
213. Cheek J, Mandelman D, Poulos TL, Dawson JH (1999) A study of the K+-site mutant of ascorbate peroxidase: Mutations of protein residues on the proximal side of the heme cause changes in iron ligation on the distal side. J Biol Inorg Chem 4:64–72
214. Bonagura CA, Sundaramoorthy M, Pappa HS, Patterson WR, Poulos TL (1996) An engineered cation site in Cytochrome *c* peroxidase alters the reactivity of the redox active tryptophan. Biochemistry 35:6107–6115
215. Barrows TP, Bhaskar B, Poulos TL (2004) Electrostatic control of the tryptophan radical in Cytochrome *c* peroxidase. Biochemistry 43:8826–8834
216. Sprangler BD, Erman JE (1986) Cytochrome *c* peroxidase compound I: formation of covalent protein crosslinks during the endogenous reduction of the active site. Biochim Biophys Acta (BBA)/Protein Struct Mol 872:155–157
217. Rittle J, Green MT (2010) Cytochrome P450 compound I: Capture, characterization, and C-H bond activation kinetics. Science 330:933–937
218. Yosca TH, Yosca TH, Rittle J, Krest CM, Onderko EL, Silakov A, Calixto JC, Behan RK, Green MT (2013) Iron(IV)hydroxide pKa and the Role of Thiolate Ligation in C–H Bond Activation by Cytochrome P450. Science 342:825–829
219. Groves JT (2014) Enzymatic C-H bond activation: Using push to get pull. Nat Chem 6:89–91
220. Wang X, Peter S, Kinne M, Hofrichter M, Groves JT (2012) Detection and kinetic characterization of a highly reactive heme-thiolate peroxygenase compound i. J Am Chem Soc 134:12897–12900
221. Wang X, Ullrich R, Hofrichter M, Groves JT (2015) Heme-thiolate ferryl of aromatic peroxygenase is basic and reactive. Proc Natl Acad Sci 201503340
222. Green MT, Dawson JH, Gray HB (2004) Oxoiron(IV) in chloroperoxidase compound II is basic: Implications for P450 chemistry. Science 304:1653–1656
223. Behan RK, Green MT (2006) On the status of ferryl protonation. J Inorg Biochem 100:448–459
224. Efimov I, Badyal SK, Metcalfe CL, Macdonald I, Gumiero A, Raven EL, Moody PCE (2011) Proton Delivery to Ferryl Heme in a Heme Peroxidase: Enzymatic Use of the Grotthuss Mechanism. J Am Chem Soc 133:15376–15383
225. Leach AR (2001) Molecular modelling: principles and applications, 2nd ed. Pearson education, Harlow, England
226. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
227. Adcock S, McCammon JA (2006) Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem Rev 106:1589–1615
228. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone

energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–15
229. Case DA, Berryman JT, Betz RM, et al (2016) AMBER 2016.
230. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–18
231. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435–447
232. Chiu J, Tillett D, Dawes IW, March PE (2008) Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. J Microbiol Methods 73:195–8
233. Lichtenstein B (2010) CloneMUTS.
234. Arslan E, Schulz H, Zufferey R, Künzler P, Thöny-Meyer L (1998) Overproduction of the *Bradyrhizobium japonicum c-type* cytochrome subunits of the cbb3 oxidase in *Escherichia coli*. Biochem Biophys Res Commun 251:744–747
235. Berry E a, Trumpower BL (1987) Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. Anal Biochem 161:1–15
236. Ensign AA, Jo I, Yildirim I, Krauss TD, Bren KL (2008) Zinc porphyrin: A fluorescent acceptor in studies of Zn-Cytochrome *c* unfolding by fluorescence resonance energy transfer. Proc Natl Acad Sci 105:10779–10784
237. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
238. Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics 7:382
239. Foloppe N, MacKerell, Jr. AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 21:86–104
240. ChemShell, a Computational Chemistry Shell, see www.chemshell.org.
241. Neese F (2012) The ORCA program system. Wiley Interdiscip Rev Comput Mol Sci 2:73–78
242. Autenrieth F, Tajkhorshid E, Baudry J, Luthey-Schulten Z (2004) Classical force field parameters for the heme prosthetic group of Cytochrome *c*. J Comput Chem 25:1613–1622
243. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
244. Shahrokh K, Orendt A, Yost GS, Cheatham TE (2012) Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. J Comput Chem 33:119–133
245. Sessions R, Ibarra A, Price J, Shoemark D, Clarke T, McIntosh-Smith S (2014) Bristol University Docking Engine (BUDE) see http://www.bris.ac.uk/biochemistry/research/bude.
246. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–68
247. Sousa da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser

interfacE. BMC Res Notes 5:367
248. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25:247–60
249. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830
250. Childs RE, Bardsley WG (1975) The steady-state kinetics of peroxidase with 2,2'-azino-di-(3-ethyl-benzthiazoline-6-sulphonic acid) as chromogen. Biochem J 145:93–103
251. D'Antonio J, Ghiladi RA (2011) Reactivity of Deoxy- and Oxyferrous Dehaloperoxidase B from Amphitrite ornata: Identification of Compound II and Its Ferrous–Hydroperoxide Precursor. Biochemistry 50:5999–6011
252. Zhao J, Lu C, Franzen S (2015) Distinct Enzyme–Substrate Interactions Revealed by Two Dimensional Kinetic Comparison between Dehaloperoxidase-Hemoglobin and Horseradish Peroxidase. J Phys Chem B 119:12828–12837
253. Klotz A V., Stegeman JJ, Walsh C (1984) An alternative 7-ethoxyresorufin o-deethylase activity assay: A continuous visible spectrophotometric method for measurement of cytochrome P-450 monooxygenase activity. Anal Biochem 140:138–145
254. Matsui T, Ozaki SI, Liong E, Phillips GN, Watanabe Y (1999) Effects of the location of distal histidine in the reaction of myoglobin with hydrogen peroxide. J Biol Chem 274:2838–2844
255. Moller K, Ottolenghi PS (1966) The oxidation of o-dianisidine by H2O2 and peroxidase at neutral pH. C R Trav Lab Carlsberg 35 16:369–389
256. Wang Q, Yang Z, Wang L, Ma M, Xu B (2007) Molecular hydrogel-immobilized enzymes exhibit superactivity and high stability in organic solvents. Chem Commun 1:1032
257. Yoo B-K, Lamarre I, Martin J-L, Andrew CR, Negrerie M (2013) Picosecond Binding of the His Ligand to Four-Coordinate Heme in Cytochrome *c* ': A One-Way Gate for Releasing Proximal NO. J Am Chem Soc 135:3248–3254
258. Raman S, Vernon R, Thompson J, et al (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins Struct Funct Bioinforma 77:89–99
259. Park H, DiMaio F, Baker D (2016) CASP11 refinement experiments with ROSETTA. Proteins Struct Funct Bioinforma 84:314–322
260. Bisswanger H (2002) Enzyme Kinetics Principles and Methods, 2nd ed. Wiley, Hoboken, NJ
261. Battistuzzi G, Borsari M, Sola M, Francia F (1997) Redox thermodynamics of the native and alkaline forms of eukaryortic and bacterial class I cytochromes c. Biochemistry 36:16247–16258
262. Murphy EJ, Metcalfe CL, Basran J, Moody PCE, Raven EL (2008) Engineering the substrate specificity and reactivity of a heme protein: creation of an ascorbate binding site in Cytochrome *c* peroxidase. Biochemistry 47:13933–13941
263. Prasain K, Nguyen TDT, Gorman MJ, Barrigan LM, Peng Z, Kanost MR, Syed LU, Li J, Zhu KY, Hua DH (2012) Redox potentials, laccase oxidation, and antilarval activities of substituted phenols. Bioorg Med Chem 20:1679–1689

264. Lefrançois P, Vajrala VSR, Arredondo IB, Goudeau B, Doneux T, Bouffier L, Arbault S (2016) Direct oxidative pathway from amplex red to resorufin revealed by in situ confocal imaging. Phys Chem Chem Phys 18:25817–25822
265. Fruk L, Müller J, Niemeyer CM (2006) Kinetic analysis of semisynthetic peroxidase enzymes containing a covalent DNA-heme add net as the cofactor. Chem - A Eur J 12:7448–7457
266. Almeida PJ, Barros AA, Rodrigues JA (1999) Determination of guaiacol at a carbon paste electrode using cathodic stripping voltammetry. Port Electrochim Acta 17:11–20
267. Barni F, Lewis SW, Berti A, Miskelly GM, Lago G (2007) Forensic application of the luminol reaction as a presumptive test for latent blood detection. Talanta 72:896–913
268. Yildiz G, Tasdoven U, Menek N (2014) Electrochemical characterization of luminol and its determination in real samples. Anal Methods 6:7809–7813
269. Al-Fartusie FSD (2011) Engineering and characterisation of novel protein covalent linkages in horseradish peroxidase (HRP): effect on structure and function. PhD Thesis. University of Sussex. doi: 10.2210/PDB2YLJ/PDB
270. Oldfield LF, Bockris JO (1951) Reversible Oxidation-Reduction Reactions of Aromatic Amines. J Phys Chem 55:1255–1274
271. Revin SB, John SA (2012) Simultaneous determination of two important dopamine metabolites at physiological pH by voltammetry. Anal Methods 4:348–352
272. Pavitt AS, Bylaska EJ, Tratnyek PG (2017) Oxidation potentials of phenols and anilines: correlation analysis of electrochemical and theoretical values. Environ Sci Process Impacts 19:339–349
273. Momeni S, Nematollahi D (2017) New insights into the electrochemical behavior of acid orange 7: Convergent paired electrochemical synthesis of new aminonaphthol derivatives. Sci Rep 7:41963
274. Jonsson M, Lind J, Reitberger T, Eriksen TE, Merényi G (1993) Redox chemistry of substituted benzenes. The one-electron reduction potentials of methoxy-substituted benzene radical cations. J Phys Chem 97:11278–11282
275. Candeias LP, Harvey PJ (1995) Lifetime and reactivity of the veratryl alcohol radical cation. Implications for lignin peroxidase catalysis. J Biol Chem 270:16745–16748
276. Joshi DK, Gold MH (1996) Oxidation of dimethoxylated aromatic compounds by lignin peroxidase from Phanerochaete chrysosporium. Eur J Biochem 237:45–57
277. Hamilton TM, Dobie-Galuska AA, Wietstock SM (1999) The o-Phenylenediamine-Horseradish Peroxidase System: Enzyme Kinetics in the General Chemistry Laboratory. J Chem Educ 76:642
278. Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, Ellisman MH, Ting AY (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat Biotechnol 30:1143–1148
279. Hosseinzadeh P, Lu Y (2016) Design and fine-tuning redox potentials of metalloproteins involved in electron transfer in bioenergetics. Biochim Biophys Acta - Bioenerg 1857:557–581
280. Bourbonnais R, Leech D, Paice MG (1998) Electrochemical analysis of the

interactions of laccase mediators with lignin model compounds. Biochim Biophys Acta 1379:381–390
281. Li C, Hoffman MZ (1999) One-Electron Redox Potentials of Phenols in Aqueous Solution. J Phys Chem B 103:6653–6656
282. Torimura M, Mochizuki M, Kano K, Ikeda T, Ueda T (1998) Mediator-assisted continuous-flow column electrolytic spectroelectrochemical technique for the measurement of protein redox potentials. Application to peroxidase. Anal Chem 70:4690–4695
283. Marková E, Kučerová P, Skopalová J, Barták P (2015) Electrochemical Oxidation of 2,4,6-Tribromophenol in Aqueous-Alcoholic Media. Electroanalysis 27:156–165
284. Gibney BR, Rabanal F, Skalicky JJ, Wand AJ, Dutton PL (1999) Iterative protein redesign. J Am Chem Soc 121:4952–4960
285. Fessl T, Watkins D, Oatley P, Allen WJ, Corey RA, Horne J, Baldwin SA, Radford SE, Collinson I, Tuma R (2018) Dynamic action of the Sec machinery during initiation, protein translocation and termination. Elife 7:e35112
286. Zahn JA, Arciero DM, Hooper AB, Dispirito AA (1996) Cytochrome *c*' of Methylococcus Capsulatus Bath. Eur J Biochem 240:684–691
287. Schagger H, Cramer WA, Vonjagow G (1994) Functional Role of Heme Ligation in Cytochrome *c*. Anal Biochem 217:220–230
288. Weiss R, Gold A, Terner J (2006) Cytochromes c': Biological models for the S = 3/2,5/2 spin-state admixture? Chem Rev 106:2550–2579
289. Lu Y, Casimiro DR, Bren KL, Richards JH, Gray HB (1993) Structurally engineered cytochromes with unusual ligand-binding properties: expression of Saccharomyces cerevisiae Met-80-->Ala iso-1-cytochrome *c*. Proc Natl Acad Sci 90:11456–11459
290. Reichlin M (1972) Hemoglobin and Myoglobin in Their Reactions with Ligands. Science 178:296–296
291. Hayakawa S, Matsumura H, Nakamura N, Yohda M, Ohno H (2013) Spectroscopic characterization of the acid-alkaline transition of a thermophilic cytochrome P450. FEBS Lett 587:94–97
292. Hayakawa S, Matsumura H, Nakamura N, Yohda M, Ohno H (2014) Identification of the rate-limiting step of the peroxygenase reactions catalyzed by the thermophilic cytochrome P450 from Sulfolobus tokodaii strain 7. FEBS J 281:1409–1416
293. Doyle WA, Blodig W, Veitch NC, Piontek K, Smith AT (1998) Two substrate interaction sites in lignin peroxidase revealed by site- directed mutagenesis. Biochemistry 37:15097–15105
294. Charizanis C, Juhnke H, Krems B, Entian K-D (1999) The oxidative stress response mediated via Pos9/Skn7 is negatively regulated by the Ras/PKA pathway in Saccharomyces cerevisiae. Mol Gen Genet MGG 261:740–752
295. Pfister TD, Gengenbach AJ, Syn S, Lu Y (2001) The role of redox-active amino acids on compound I stability, substrate oxidation, and protein cross-linking in yeast Cytochrome *c* peroxidase. Biochemistry 40:14942–14951
296. Lardinois OM, Maltby DA, Medzihradszky KF, Ortiz De Montellano PR, Tomer KB, Mason RP, Deterding LJ (2009) Spin scavenging analysis of myoglobin

protein-centered radicals using stable nitroxide radicals: Characterization of oxoammonium cation-induced modifications. Chem Res Toxicol 22:1034–1049
297. Chateauneuf J, Lusztyk J, Ingold KU (1988) Absolute Rate Constants for the Reactions of Some Carbon-Centered Radicals with 2,2,6,6-Tetramethylpiperidine-N-oxyl. J Org Chem 53:1629–1632
298. Wojdyr M (2010) Fityk: A general-purpose peak fitting program. J Appl Crystallogr 43:1126–1128
299. Holzwarth G, Doty P (1965) The Ultraviolet Circular Dichroism of Polypeptides. J Am Chem Soc 87:218–228
300. Berglund GI, Carlsson GH, Smith AT, Szoke H, Henriksen A, Hajdu J (2002) The catalytic pathway of horseradish peroxidase at high resolution. Lett to Nat 417:463–468
301. Moody PCE, Raven EL (2018) The Nature and Reactivity of Ferryl Heme in Compounds I and II. Acc Chem Res 51:427–435
302. Wright PJ, English AM (2003) Scavenging with TEMPO. to identify peptide- and protein-based radicals by mass spectrometry: Advantages of spin scavenging over spin trapping. J Am Chem Soc 125:8655–8665
303. Dunford HB, Stillman JS (1976) On the function and mechanism of action of peroxidases. Coord Chem Rev 19:187–251
304. Djukic B, Dube PA, Razavi F, Seda T, Jenkins HA, Britten JF, Lemaire MT (2009) Preparation and Magnetic Properties of Iron(3+) Spin-Crossover Complexes Bearing a Thiophene Substituent: Toward Multifunctional Metallopolymers. Inorg Chem 48:699–707
305. Pond AE, Bruce GS, English AM, Sono M, Dawson JH (1998) Spectroscopic study of the compound ES and the oxoferryl compound II states of Cytochrome *c* peroxidase: comparison with the compound II of horseradish peroxidase. Inorganica Chim Acta 275–276:250–255
306. Fishel LA, Farnum MF, Miller MA, Kraut J, Mauro JM, Liu Y, Tan X ling, Scholes CP (1991) Compound I Radical in Site-Directed Mutants of Cytochrome *c* Peroxidase As Probed by Electron Paramagnetic Resonance and Electron-Nuclear Double Resonance. Biochemistry 30:1986–1996
307. Svistunenko DA, Cooper CE (2004) A New Method of Identifying the Site of Tyrosyl Radicals in Proteins. Biophys J 87:582–595
308. Thompson MK, Franzen S, Ghiladi RA, Reeder BJ, Svistunenko DA (2010) Compound ES of Dehaloperoxidase Decays via Two Distal Histidine. J Am Chem Soc 132:17501–17510
309. Hüttemann M, Pecina P, Rainbolt M, Sanderson TH, Kagan VE, Samavati L, Doan JW, Lee I (2011) The multiple functions of Cytochrome *c* and their regulation in life and death decisions of the mammalian cell: From respiration to apoptosis. Mitochondrion 11:369–381
310. Kagan VE, Tyurin VA, Jiang J, et al (2005) Cytochrome *c* acts as a cardiolipin oxygenase required for release of proapoptotic factors. Nat Chem Biol 1:223–232
311. Hanske J, Toffey JR, Morenz AM, Bonilla AJ, Schiavoni KH, Pletneva E V. (2012) Conformational properties of cardiolipin-bound Cytochrome *c*. Proc Natl Acad Sci 109:125–130
312. Rajagopal BS, Edzuma AN, Hough MA, et al (2013) The hydrogen-peroxide-

induced radical behaviour in human Cytochrome *c* –phospholipid complexes: implications for the enhanced pro-apoptotic activity of the G41S mutant. Biochem J 456:441–452
313. Ziegler T (1991) Approximate density functional theory as a practical tool in molecular energetics and dynamics. Chem Rev 91:651–667
314. Kohn W, Sham LJ (1965) Self-Consistent Equations Including Exchange and Correlation Effects. Phys Rev 140:A1133–A1138
315. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104
316. Lonsdale R, Harvey JN, Mulholland AJ (2012) Effects of dispersion in density functional based quantum mechanical/molecular mechanical calculations on cytochrome P450 catalyzed reactions. J Chem Theory Comput 8:4637–4645
317. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98:5648–5652
318. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37:785–789
319. Perdew JP, Wang Y (1992) Accurate and Simple Analytic Representation of the Electron-Gas Correlation-Energy. Phys Rev B 45:13244–13249
320. Perdew JP (1991) Electronic structure of solids' 91. Akademie Verlag, Berlin
321. Jeschek M, Reuter R, Heinisch T, Trindler C, Klehr J, Panke S, Ward TR (2016) Directed evolution of artificial metalloenzymes for *in vivo* metathesis. Nature 537:661–665
322. Hay PJ, Wadt WR (1985) *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. J Chem Phys 82:299–310
323. Pople JA, Nesbet RK (1954) Self-Consistent Orbitals for Radicals. J Chem Phys 22:571–572
324. Lonsdale R, Oláh J, Mulholland AJ, Harvey JN (2011) Does compound I vary significantly between isoforms of cytochrome P450? J Am Chem Soc 133:15464–74
325. Field MJ, Bash PA, Karplus M (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11:700–733
326. Mulholland AJ (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. Drug Discov Today 10:1393–1402
327. Amara P, Field MJ (2003) Evaluation of an *ab initio* quantum mechanical/molecular mechanical hybrid-potential link-atom method. Theor Chem Accounts Theory, Comput Model (Theoretica Chim Acta) 109:43–52
328. Bakowies D, Thiel W (1996) Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. J Phys Chem 100:10580–10594
329. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107:3902–3909
330. Derat E, Cohen S, Shaik S, Altun A, Thiel W (2005) Principal active species of horseradish peroxidase, compound I: A hybrid quantum mechanical/molecular

mechanical study. J Am Chem Soc 127:13611–13621
331. Bathelt C, Mulholland AJ, Harvey JN (2005) QM/MM studies of the electronic structure of the compound I intermediate in Cytochrome *c* peroxidase and ascorbate peroxidase. Dalton Trans 3470–3476
332. Lonsdale R, Rouse SL, Sansom MSP, Mulholland AJ (2014) A multiscale approach to modelling drug metabolism by membrane-bound cytochrome P450 enzymes. PLoS Comput Biol 10:e1003714
333. Green AP, Hayashi T, Mittl PRE, Hilvert D (2016) A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. J Am Chem Soc 138:11344–11352
334. Goodin DB, McRee DE (1993) The Asp-His-Fe Triad of Cytochrome-C Peroxidase Controls the Reduction Potential, Electronic-Structure, and Coupling of the Tryptophan Free-Radical to the Heme. Biochemistry 32:3313–3324
335. Goodin DB (1996) When an amide is more like histidine than imidazole: The role of axial ligands in heme catalysis. J Biol Inorg Chem 1:360–363
336. Raven EL (2003) Understanding functional diversity and substrate specificity in haem peroxidases: What can we learn from ascorbate peroxidase? Nat Prod Rep 20:367–381
337. Rydberg P, Sigfridsson E, Ryde U (2004) On the role of the axial ligand in heme proteins: A theoretical study. J Biol Inorg Chem 9:203–223
338. Ponder JW, Case DA (2003) Force Fields for Protein Simulations. In: Adv. Protein Chem. pp 27–85
339. MacKerell AD, Jr, Bashford D, et al (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B 102:3586–3616
340. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–12
341. Martins-Costa MTC, Ruiz-López MF (2007) Molecular dynamics of hydrogen peroxide in liquid water using a combined quantum/classical force field. Chem Phys 332:341–347
342. Zoete V, Cuendet MA, Grosdidier A, Michielin O (2011) SwissParam: A fast force field generation tool for small organic molecules. J Comput Chem 32:2359–2368
343. Sherwood P, de Vries AH, Guest MF, et al (2003) QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol Struct THEOCHEM 632:1–28
344. Mills JEJ (1996) Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. J Comput Aided Mol Des 10:607–622

Appendix

A. Chapter 1

These are no appendices for Chapter 1.

B. Chapter 2

B.1 Rosetta ab initio

----Flags---

-abrelax
-abinitio:relax
-abinitio::increase_cycles 10
-relax:fast
-relax:jump_move
-score:weights score13_env_hb
-abinitio:rg_reweight 0.5
-abinitio:rsd_wt_helix 0.5
-abinitio:rsd_wt_loop 0.5
-jumps:residue_pair_jump_file ../input/$molname.residue_pair_jump_cst
-use_filters true
-psipred_ss2 ../input/$molname.psipred_ss2
-in:file:fasta ../input/$molname.fasta
-in:file:frag3 ../input/$molname_03_05.200_v1_3
-in:file:frag9 ../input/$molname_09_05.200_v1_3
-constraints:cst_file ../input/$molname.cen_cst
-constraints:cst_weight 1.0
-constraints:cst_fa_file ../input/$molname.fa_cst
-constraints:cst_fa_weight 1.0
-out:nstruct 1
-out:pdb
-extra_res_cen ../input/heme-amber.cen.params
-extra_res_fa ../input/heme-amber.fa.params

--- C45.residue_pair_jump_cst---

BEGIN
jump_def: 114 139 138 138
aa: CYZ HEM
cst_atoms: SG CB CA CAB C3B C4B
jump_atoms: N CA C CAB C3B C4B
disAB: 1.80
angleA: 114.5
angleB: 180
END

B.2 Amber input files

```
   ----- minsolv.i -----
   Minimization of solvent+Hs with
restraints for protein heavy atoms
   &cntrl
    imin=1, maxcyc=50, ntmin=1,
    ntpr=10,
    cut=12.0,
    ntr=1, restraint_wt=100.0,
    restraintmask='!:WAT,SOD,CLA &
!@H='
   /

   ----- mdsolv.i -----
   MD heating of solvent with
restraints for solute
   &cntrl
    imin=0, irest=0, ntx=1,
    nstlim=25000, dt=0.002,
    ntpr=500, ntwx=5000, ioutfm=1,
    ntb=2, cut=12.0,
    ntc=2, ntf=2,
    ntt=1, tautp=1.0, tempi=50.0,
temp0=300.0,
    ntp=1, taup=1.0, pres0=1.0,
    ntr=1, restraint_wt=25.0,
restraintmask='!:WAT,SOD,CLA'
   &end
   &ewald
    skinnb=3.0
   &end

   ----- min.i -----
   Minimization of system
   &cntrl
    imin=1, maxcyc=300, ntmin=1,
    ntpr=50,
    cut=12,
   &end
```

```
   ----- heat.i -----
   MD quick heating of system
   &cntrl
    imin=0, irest=0, ntx=1, ig=-1,
    nstlim=10000, dt=0.002,
    ntpr=100, ntwx=5000, ioutfm=1,
iwrap=0,
    ntb=1, cut=12.0,
    ntc=2, ntf=2,
    ntt=1, tautp=1.0, tempi=25.0,
temp0=300.0,
   /


   ----- eq_ntp.i -----
   200ps NPT MD equilibration of
system
   &cntrl
    imin=0, irest=1, ntx=5,
    nstlim=100000, dt=0.002,
    ntpr=1000, ntwx=50000, ioutfm=1,
iwrap=0,
    ntc=2, ntf=2,
    ntb=2, cut=12.0,
    ntp=1, pres0=1.0, taup=1.0,
    ntt=1, tempi=300.0, tautp=1.0,
   /

   ----- md.i -----
   10ns NPT MD production simulation
   &cntrl
    imin=0, irest=1, ntx=5,
    nstlim=5000000, dt=0.002,
    ntpr=10000, ntwx=50000,
ioutfm=1, iwrap=1,
    ntc=2, ntf=2,
    ntb=2, cut=12.0,
    ntp=1, pres0=1.0, taup=1.0,
    ntt=1, tempi=300.0, tautp=0.1,
   /
```

----- RES.dis ------

```
&rst
 ixpk= 0, nxpk= 0, iat= 851,4638, r1= 1.30, r2=
1.80, r3= 5.70, r4= 6.10,
    rk2=0.0, rk3=10.0,
&end
&rst
 ixpk= 0, nxpk= 0, iat= 3662,2727, r1= 1.10, r2= 1.20, r3= 2.50, r4= 3.50,
    rk2=0.0, rk3=5.0,
&end
&rst
 ixpk= 0, nxpk= 0, iat= 3663,2967, r1= 1.10, r2= 1.20, r3= 2.50, r4= 3.50,
    rk2=0.0, rk3=5.0,
&end
```

B.3 ChemShell input script example

```
# QM/MM calculation of ferric_CcP from amber MD cluster
# JJ 12.03.17
# Using dl-find
# getting $molname
source ./molname_vmd.dat
# getting $qmatoms
source ./qm_atoms.dat
# getting $active amd $frozen atoms
source ./frozen_atoms.dat
source ./active_atoms.dat
set file_name "$molname-qmmm3C"
puts stdout "starting $file_name"
set Nproc "4"
set resxyz "vmd-out"
puts stdout "---- This script is for calculatons without the propionates ----"
set charge1 [ expr $Charge + 2 ]
puts stdout "charge changed from $Charge to $charge1"
set Charge $charge1
puts stdout "mult: $Mult charge: $Charge"
# Define topology and parameter files
set top /panfs/panasas01/chem/jj1134/chemshell-
charmm/top_all27_prot_na_heme.rtf
set par /panfs/panasas01/chem/jj1134/chemshell-
charmm/par_all27_prot_na_heme.prm
read_xplor_psf atom_types= types atom_charges= charges atom_masses= mass file=
$file_name.psf
# save data
set fp [ open save.chm w]
puts $fp "set types   [ list $types ]"
puts $fp "set charges [ list $charges]"
# Read starting coordinates from PDB file
puts stdout "readining in coordinates from $file_name.pdb..."
read_pdb coords=start.c file=$file_name.pdb
read_pdb coords=start-org.c file=$file_name.pdb
puts stdout "done"
#### Read starting coordinates from VMD xyz file
puts stdout "readining in coordinates from $resxyz.xyz..."
read_xyz coords=start.c file=$resxyz.xyz
puts stdout "done"
# Obtain residue information (for dl-find) from PDB file
puts stdout "setting pdb_residues..."
set pdb_residues [ pdb_to_res "$file_name.pdb"  ]
puts stdout "done"
# Sourcing the new ORCA interface
source  /panfs/panasas01/chem/jj1134/SOFTWARE/orca-chemsh.tcl
# ORCA block settings are specified here.
set orcasimpleinput "! B3LYP 6-31G Opt LooseSCF D3"
set orcablocks "
%basis
  # Dummy basis, will be overridden
  NewGTO fe \"LANL2DZ\" end
  NewECP fe \"LANL2\" end
end

%output
Print\[ P_Basis \] 2
Print\[ P_MOs \] 1
end

%plots
dim1 100
dim2 100
dim3 100
Format Gaussian_Cube
SpinDens(\"spindensity\");
ElDens(\"electrondensity\");
end
```

```
%scf
  HFTyp uhf
  MaxIter 400
end
"
set fp [ open pes.txt w ]
# Loop over the potential energy surface
set rc "$file_name"
# Geometry Optimization
dl-find coords=start.c \
     coordinates= hdlc \
     residues= $pdb_residues \
     frozen= $frozen \
     active_atoms= $active \
     optimiser=lbfgs \
     maxcycle= 500 \
     dump = 10 \
     theory=hybrid : [ list \
       coupling=shift \
       atom_charges= $charges \
       qm_theory=orca : [ list  \
          executable=/panfs/panasas01/chem/jj1134/SOFTWARE/orca_4_0_0_2_linux_
x86-64/orca \
          orcasimpleinput= $orcasimpleinput \
          orcablocks= $orcablocks \
          nproc=$Nproc \
          charge=$Charge \
          mult=$Mult ] \
       qm_region= $qmatoms \
       mm_theory=dl_poly : [ list \
          mxexcl=176 \
          conn=start-org.c \
          exact_srf=yes \
          use_pairlist=no \
          cutoff=1000 \
          scale14 = { 1.0 1.0 } \
          atom_types= $types \
          atom_charges= $charges \
          use_charmm_psf=yes \
          charmm_psf_file=$file_name.psf \
          charmm_parameter_file=$par \
          charmm_mass_file=$top ] ] \
     list_option=full \
     result= result.c

set final_energy [ get_matrix_element matrix=dl-find.energy indices= {0 0 } ]
puts "Reaction coordinate: $rc Energy: $final_energy"
puts $fp "$rc $final_energy"
set tcl_precision 12
# Write trajectory
write_xyz file=tmp.xyz coords=result.c
exec cat tmp.xyz >> pes.xyz
# Write pdb for this step
write_pdb file=rc_$rc.pdb coords=result.c
close $fp
times
```

C. Chapter 3

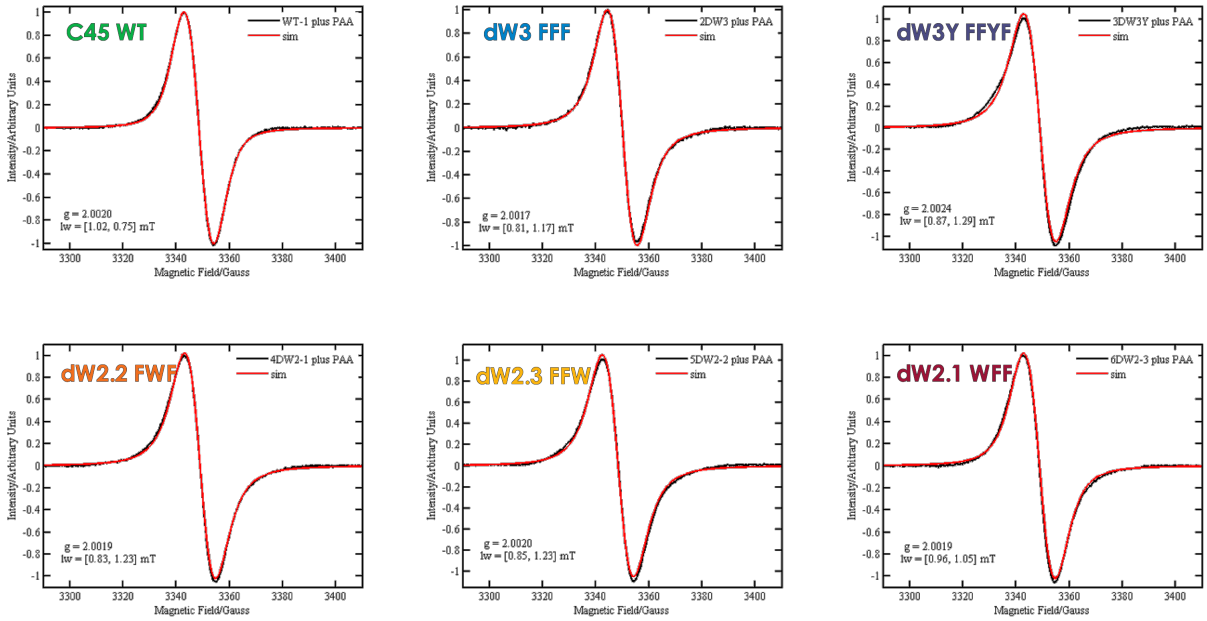


Figure C.1 Experimental (black) and simulated (red) results for EPR radical signals. The g-tensor value and spin-Hamiltonian parameters used to model the EPR signal is also given as an inset in each figure. The naming convention between the experimental samples of the dW2.1-3 mutants and the names used in the thesis is not consistent.

D. Chapter 4

Structure			Distance (Å)								
			NS 4CVI	110-ferric-CcP-HID	120-ferric-CcP-HIP	NS 4CVJ	210-c1-CcP-HIP	220-c1-CcP-HID	NS 5JPR	310-c2-APX-HIP	320-c2-APX-HID
Cycles			0	89	77	0	79	75	0	86	147
HEM:**FE**	-	pHSD:**NE2**	**2.02**	**2.10**	**2.15**	**2.12**	**2.13**	**2.13**	**1.97**	**2.04**	**2.06**
pASP:**OD2**	-	pHSD:**HD1**	2.06	1.73	2.11	2.11	2.51	3.44	2.32	3.46	2.63
pASP:**OD2**	-	pTRP:**NE1**	2.65	2.80	2.70	2.69	2.71	2.72	2.73	2.71	2.70
pTRP:**HE1**	-	pTRP:**NE1**	0.87	1.04	1.70	0.86	1.72	1.72	0.86	1.71	1.70
HEM:**FE**	-	dHIS:**NE2**	5.73	6.22	5.78	5.21	5.76	5.53	5.33	5.91	5.02
HEM:**FE**	-	dTRP:**NE1**	3.72	4.34	3.61	3.15	2.92	3.07	3.26	3.02	3.67
HEM:**FE**	-	dARG:**NH2**	7.16	5.79	6.29	4.33	5.00	4.81	7.67	4.97	5.34
WAT:**O**	-	HEM:**FE**	2.65	3.83	3.36	3.92	3.83	3.95	4.15	3.92	4.00
WAT:**O**	-	dHSD:**NE2**	3.10	2.67	3.03	2.63	2.77	2.81	2.81	2.75	3.84
WAT:**O**	-	dTRP:**NE1**	2.32	2.87	3.51	2.68	3.70	3.81	2.75	3.42	4.11
WAT:**O**	-	dARG:**NH2**	5.06	2.87	4.09	3.44	2.88	2.73	6.65	2.90	2.83
HEM:**FE**	-	HEM:OE				**1.64**	**1.67**	**1.65**	**1.88**	**1.89**	**1.86**
HEM:**OE**	-	WAT:**O**				3.06	2.63	2.75	2.96	2.60	2.68
HEM:**OE**	-	HEM:**HO1**							0.99	0.97	0.96

Table D.1 QM cluster model distance information for QM1-Di result with solvation field.

Species	Difference in Energy (kcal/mol)					
	CcP	**CcP-SUB**	**APX**	**APX-SUB**	**APX2**	**APX2-SUB**
ΔEnF, Ferric, P→D	252	-57	254	527	533	394
ΔEn1, Cpd I, P→D	631		389	508	180	328
ΔEn2, Cpd II, P→D	788		655	289	750	403

Table D.2 Calculated QM/MM energy differences between the dHIS pairs.

E. Structural mutants

This appendix contains preliminary data on structural mutants of *c*-type maquettes. CXH is a *c*-type maquette that is an extended version of C45. It has an additional heptad on each helix (Figure E.1). CCC is a coiled-coil-C45 chimera the helices at the *c*-type heme binding end of the bundle are the same sequence as C45. The remaining sequence comes from an anti-parallel coiled-coil. Genes were designed and optimised before being produced (Eurofins, UK). They were then cloned into the same expression vector as C45 (Section 2.3.1).

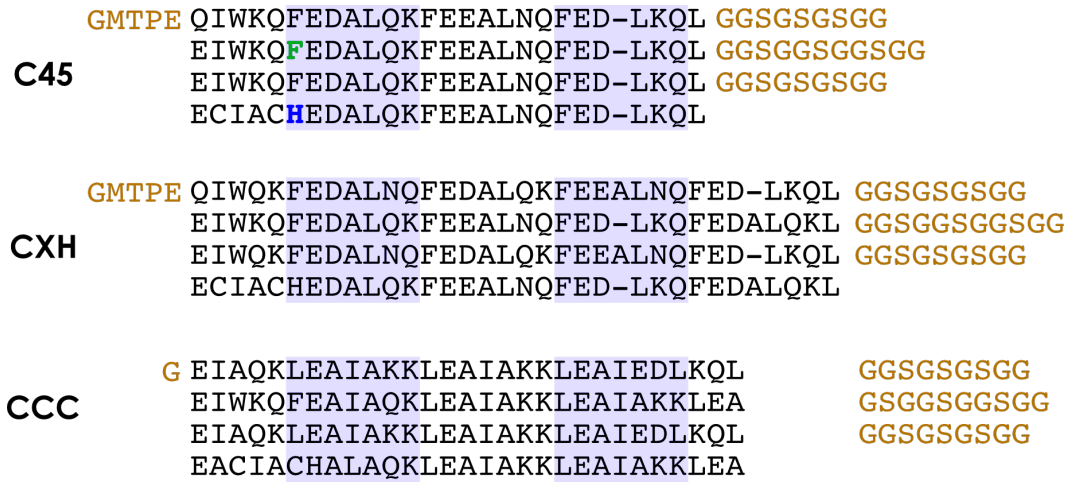


Figure E.1 Protein sequences of new structural mutants.

E.1 CXH cloning expression purification

Only CXH was expressed and purified. The expression and purification of CXH was conducted using the same method as C45 (seen in Section 2.2.2). Figure E.2 shows the size exclusion chromatography (SEC) trace and SDS-PAGE results for the purification of CXH. The SEC trace shows that CXH is not in a homogeneous conformation at pH 8.6. There is a secondary band seen in the SDS-PAGE results close to the same weight as CXH that may correspond to holoenzyme without heme bound.

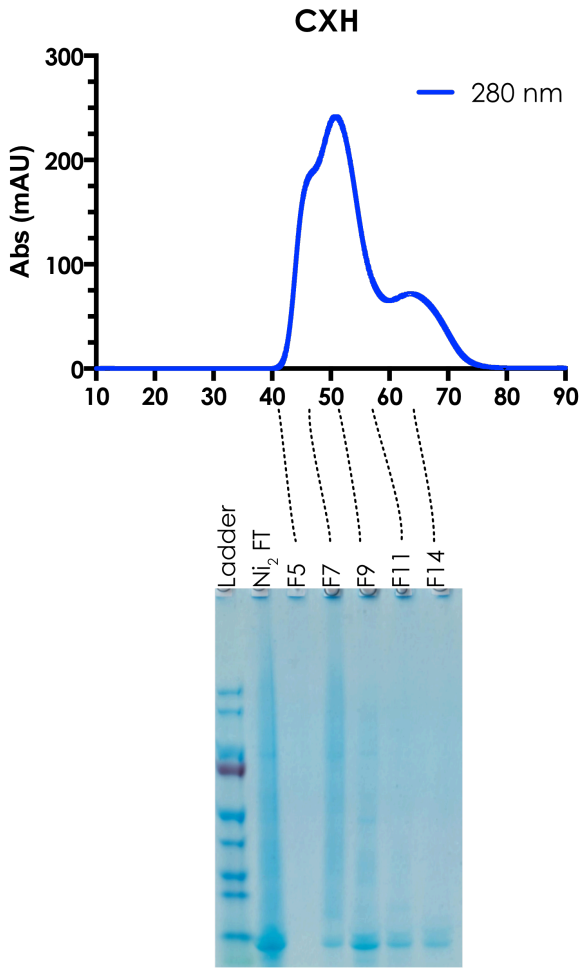


Figure E.2 ***SEC and SDS-PAGE results for CXH.***

The 280 nm SEC trace is shown with the x-axis showing the accumulated volume (ml). Below the SEC trace is the SDS-PAGE results of the purification. The bands (not including the ladder) start with "Ni2 FT" lane which is the elution from the first nickel affinity column after the cleavage of the his-tag. The remaining bands are fractions from the SEC purification of the flow-through from the second nickel column.

E.2 UV-visible Spectroscopy

The ratio of protein (280 nm) to Soret λ_{max} is reduced in CXH. It is 1:2 rather than 1:3 seen in C45. This is likely a result of a population of holoenzyme in the CXH sample that failed to incorporate *c*-type heme.

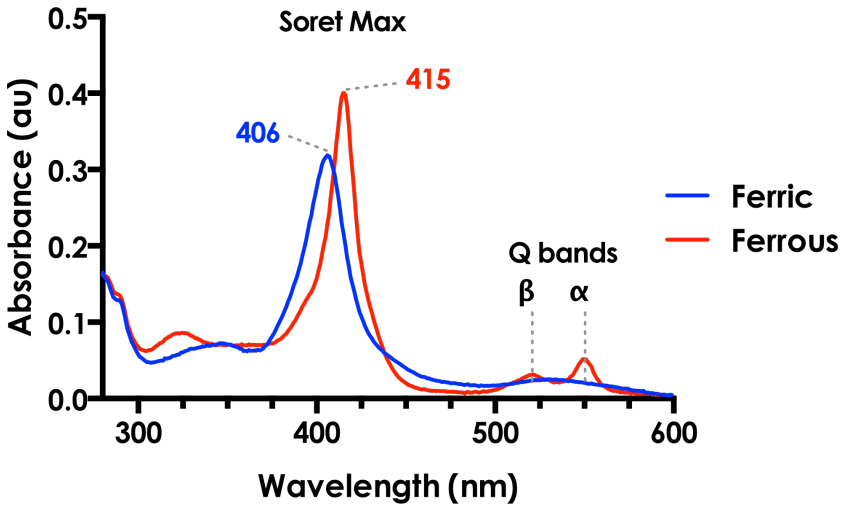


Figure E.3 UV-visible spectra of CXH in the ferric and ferrous form.

E.3 CD

CD spectrum of CXH shows the typical absorbance's for α-helical proteins. These are negative bands at 222 nm and 208 nm and a positive band at 193 nm. The melting temperature has increased in comparison to C45. CXH has a Tm of 88 °C where as C45 has a Tm of 84 °C.

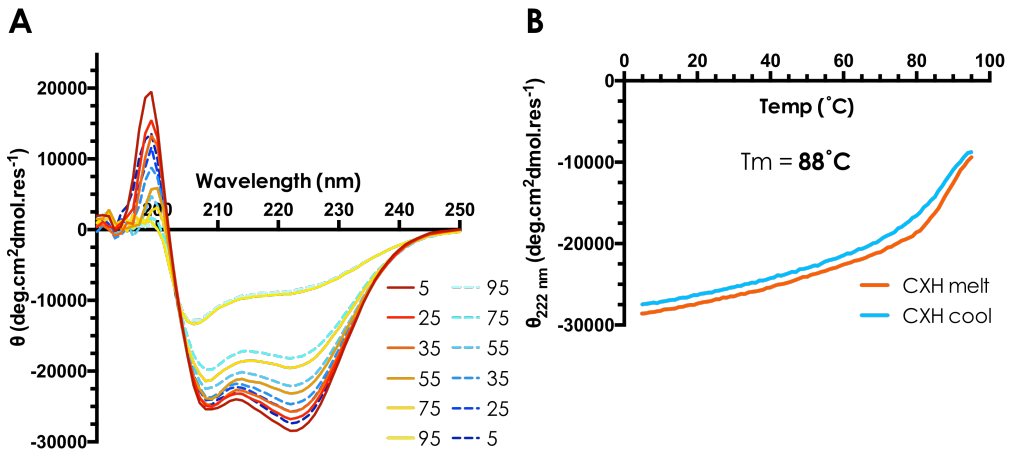


Figure E.4 CD results for CXH. ***(A)*** *CD spectrum for CXH. Scans of melts for 5-95°C are displayed as solid lines going from red to yellow. Scans of cooling experiments from 95-5°C are displayed as dashed lines from light to dark blue.* ***(B)*** *the mean residue ellipticity (θ) at 222 nm during both melt and cool.*

E.4 Electrochemistry

Optically transparent thin-layer electrochemistry (OTTLE) was performed by the same method seen in Section 3.2.7. CXH has a EM only 6 mV more negative that C45 (CXH = 180.2 C45 = 174 mV vs NHE).

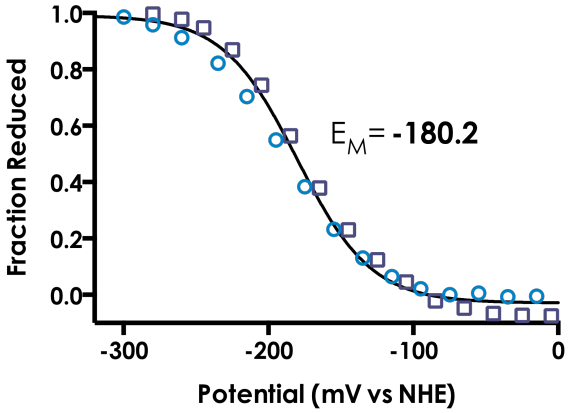


Figure E.5 ***OTTLE results.*** *The fraction of reduced protein has been fitted to a single electron Nernst equation. The midpoint potentials (EM) are shown on each graph. The values taken during the reduction are shown in circles and the subsequent oxidation values are shown as squares.*

E.5 Kinetic assays

The concentration of CXH was estimated using the *c*-type Soret λ_{max} extinction coefficient determined for C45. This will have led to and underestimate in the concentration of total protein used in all the preliminary assays. However, it may be a good estimate of the amount of protein with *c*-type heme incorporated, as the concentration measurements was determined using the heme absorbance (Soret λ_{max}). The same methodology for steady state kinetics with ABTS was used as described in Section 2.2.12. CXH has very similar steady-state kinetic properties to C45. Table E.5.1 gives a comparison of the Michaelis–Menten fit values at rate limiting H_2O_2 concentration.

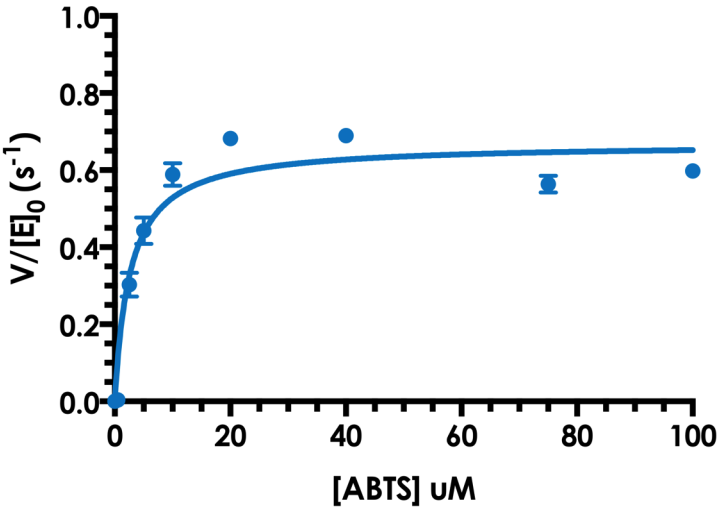


Figure E.6 Steady-state kinetics of CXH with ABTS. The error bars are the standard deviation.

Enzyme	Substrate	k_{cat}* (s^{-1})	K_M (μM)	k_{cat}/K_M ($M^{-1}s^{-1}$)	Condition
C45 WT	ABTS	2.0*± 0.028	3.1± 0.21	6.4± 0.43x10 ^{5}	pH 8.6
CXH	ABTS	0.67*± 0.0002	2.7± 0.38	2.5± 0.36x10 ^{5}	pH 8.6

Table E.5.1 Steady-state kinetics curve fitting values. k_{cat}* values are displayed as the results are a rate limiting H_2O_2 concentration (100 μM).